Novembre
2024

numéro
24

1024 Bulletin de la Société informatique de France

COMITÉ DE RÉDACTION

DENIS PALLEZ, *Université Côte d'Azur, rédacteur en chef*
Contact : 1024@societe-informatique-de-france.fr

SYLVIE ALAYRANGUES
Université de Poitiers

OLIVIER BAUDON
Université de Bordeaux

YVES BERTRAND
Université de Poitiers

JEAN-PAUL DELAHAYE
Université de Lille

ISABELLE DEBLED-RENNESSON
Université de Lorraine

GIUSEPPE DI MOLFETTA
Université Aix-Marseille

MARIE DUFLOT-KREMER
Université de Lorraine

FRÉDÉRIC HAVET
CNRS, Université Côte d'Azur

MATTHIEU LATAPY
CNRS, Sorbonne Université

BAPTISTE MÉLÈS
CNRS, Université de Lorraine

BRUNO MERMET
Université du Havre

PATRICE NAUDIN
Université de Poitiers

NICOLAS PASSAT
Université de Reims Champagne-Ardenne

MICHEL RAYNAL
Université de Rennes

LAURENT RÉVEILLÈRE
Université de Bordeaux

NANCY RODRIGUEZ
Université de Montpellier

FLORENCE SÈDES
Université de Toulouse

LAURENT SIGNAC
Université de Poitiers

Cette œuvre est mise à disposition sous licence Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0. Pour voir une copie de cette licence, visitez http://creativecommons.org/licenses/by-nc-nd/4.0/deed.fr *ou écrivez à Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.*

SOCIÉTÉ INFORMATIQUE DE FRANCE
Institut Henri Poincaré, 11 rue Pierre et Marie Curie, 75231 Paris Cedex 05

Prix public : 32 € (adhérents SIF : −30%)

Directeur de la publication : Yves Bertrand ISSN : 2270-1419

Couverture réalisée par l'agence graindecel, https://www.bauwensceline.com.
Composition réalisée par C&F éditions, https://cfeditions.com.

Éditorial

Le bulletin que vous avez entre les mains n'est pas tout à fait le bulletin habituel… en effet, ce numéro 24 est le premier résultat concret du travail accompli depuis plusieurs mois par l'éditeur C&F éditions en collaboration avec la SIF pour concevoir une nouvelle chaîne de fabrication de 1024, plus professionnelle. Vous connaissez probablement déjà C&F éditions si vous avez reçu les deux livres : « Qui a hacké Garoutzia ? » et « Les entretiens de binaire », ouvrages qui ont été distribués avec le bulletin 1024 ces deux dernières années. Dans le numéro 20 de ce bulletin[1], nous annoncions déjà le projet de faire réaliser 1024 par un éditeur professionnel, afin de pouvoir concentrer les efforts du comité de rédaction sur des tâches plus conformes au rôle traditionnel qu'il joue habituellement dans la vie d'un journal (appel aux communications, sélection des articles, etc.).

De plus, ce nouveau procédé de fabrication de notre bulletin devrait permettre, dans un avenir très proche, de vous proposer de nouveaux services comme la lecture de chaque article directement dans votre navigateur Internet (y compris sur votre mobile) ou sur votre liseuse préférée. Je tiens — en tant que rédacteur en chef de 1024 — à remercier la SIF pour l'investissement financier qu'elle a consenti pour que nous arrivions à ce résultat.

L'article « Un nouveau moteur pour 1024 », inclus dans ce bulletin, écrit par les deux fondateurs de C&F éditions, Hervé Le Crosnier et Nicolas Taffin, explique dans les détails tous les défis qu'il a fallu relever pour arriver à ce résultat.

Comme ce bulletin est *votre* bulletin, nous serions heureux d'avoir votre avis sur cette nouvelle formule : présentation interne des articles, structure des pages, etc. tout élément qui vous semble important dans un bulletin comme le nôtre.

Comme une bonne nouvelle arrive rarement seule, nous avons également le plaisir de vous proposer un tout nouveau défi intitulé c0d1ngUP (lire *Coding up*).

1. https://doi.org/10.48556/SIF.1024.20.3.

DOI:10.48556/SIF.1024.24.3

Le projet d'insertion dans 1024 d'un challenge de programmation remonte à plusieurs années, mais nous n'avions pas réussi à le concrétiser jusque-là. Eh bien ! c'est maintenant chose faite avec cette nouvelle rubrique récurrente. Au même titre que les énigmes algorithmiques ou combinatoires proposées par Jean-Paul Delahaye — qui a, de nouveau, inspiré notre dessinateur attitré, Frédéric Havet, avec un décalage toujours aussi subtil et plein d'humour — un défi de programmation sera proposé dans chaque numéro avec des pistes de solutions dans le numéro suivant. L'un des points forts de ce casse-tête de programmation est qu'il n'impose pas de langage de programmation particulier ni même de méthode particulière mais, au contraire, laisse les « compétiteurs » libres de leur approche des problèmes. Longue vie à ce nouveau défi et merci aux membres de l'équipe pilote de faire confiance à 1024 pour diffuser leur travail.

Ce numéro fait encore la part belle à l'enseignement, qui occupe beaucoup d'entre nous, et est l'un des axes prioritaires de l'action de notre société savante. Plusieurs éléments de l'enseignement de l'informatique (essentiellement dans le secondaire) sont abordés, parfois de manière un peu polémique. Tout d'abord, un premier article fait un *bref* compte rendu du concours de l'agrégation 2023 (seconde édition de ce concours) : un résumé du rapport du jury, avec les plus importants commentaires, recommandations aux futurs candidats, etc.

Un second article est dédié à une étude en profondeur du fonctionnement de la plateforme Parcoursup, de l'algorithme qui en constitue le fondement, et sur l'utilisation qui est faite de cet algorithme. L'article étudie en détail le processus d'affectation, à la fois du côté de l'administration mais aussi du point de vue des candidats… qui attendent, attendent… Et l'article met l'accent sur la possibilité pour les candidats de définir des priorités sur les établissements dans lesquels ils envisagent d'entamer leurs études supérieures (comme c'était le cas d'Admission postbac avant l'arrivée de Parcoursup).

L'article « Nouveau lycée : plus spécialisé que moi, tu meurs ! » est une étude sur les spécialités que nos lycéens doivent désormais choisir en classe de première et de terminale et milite pour revenir à une situation dans laquelle les lycéens pourraient conserver en terminale les trois spécialités qu'ils ont choisies en première. Ceci permettrait probablement de juguler la désaffection dont sont victimes certaines spécialités au passage de la première à la terminale au profit, par exemple, de la spécialité « Mathématiques ». NSI, également victime de cette réduction du nombre de spécialités, est passée au peigne fin des statistiques dans le dernier article « La spécialité NSI… péril en la demeure ! ». Choix de spécialités et désaffection à l'égard de NSI sont également replacés dans le contexte d'une déféminisation des sciences en général et de l'informatique en particulier.

Deux articles nous parlent de la science et de l'informatique dans la société. Le collège des sociétés savantes — créé il y a presque quatre ans — œuvre pour peser dans le débat public et associer les scientifiques aux décisions des

politiques relevant du domaine de la Science ; son président, Patrick Lemaire, nous explique ce collège sous tous les aspects de sa vie. Même si c'est un problème plus spécifiquement lié à l'informatique, l'ouverture des codes sources des applications de l'administration de l'État participe également de cette problématique et contribue à la transparence et au contrôle de politiques publiques en la matière ; l'article sur la « citoyenneté informatique » donne quelques éléments de réflexion sur ce sujet.

Le côté sciences informatiques n'est pas en reste dans ce numéro. On peut y lire le second article (d'une série de trois) consacré à la 5G, et aux problèmes environnementaux auquels l'implémentation de cette nouvelle technologie est confrontée, ce qui incite à relativiser grandement l'enthousiasme que l'on peut avoir vis-à-vis de la 5G. Un article, de l'un des deux récipiendaires du prix Lovelace-Babbage 2023 (délivré par l'Académie des sciences et la SIF), sur les interfaces cerveau-ordinateur est consacré à ces technologies qui restent encore futuristes : piloter des processus physiques directement par la pensée, cela semble magique, eh bien, on y vient tout doucement, et il faut lire cet article.

Le prix Gilles Kahn est délivré chaque année par la SIF et le lauréat du prix ainsi qu'une des deux lauréates d'un accessit nous présentent leurs travaux dans deux articles. Le premier nous parle de la représentativité de certaines données issues du calcul dans de grandes collections de données. La seconde nous explique comment l'on peut détecter qu'une image JPEG a été retouchée (je simplifie outrageusement).

Un article nous propose ensuite une étude sur la cryptanalyse par des moyens physiques dans le contexte du chiffrement symétrique. Un autre article décrit une application d'analyse (par graphes de connaissance) de données provenant de Wikidata ; des techniques analogues, appliquées aux grandes quantités de données de Wikidata pourraient-elle aider à retrouver des œuvres d'art pillées ?

Que ce soit dans le panel d'articles de ce numéro ou bien dans les exposés au congrès de la SIF de juin dernier —, les études et techniques liées de près ou de loin à l'IA sont de plus en plus présentes dans nos colonnes… c'est une bonne raison de jeter un coup d'œil au livre qu'a lu pour nous Bruno Mermet : un point de vue intéressant sur ce que l'on nomme « intelligence artificielle ».

Enfin, on trouve dans ce numéro deux articles sur deux aspects (très différents) de l'histoire de l'informatique. Tout d'abord un petit mémo pour fêter l'anniveraire de la naissance du langage Basic, né au milieu des années soixante, en même temps que la souris (la souris informatique, du PARC de Xerox). Et puis, un article consacré à une image, et à une femme : l'image est la photo que tout informaticien a pu voir dans les articles publiés par ses collègues spécialistes d'analyse d'images, et la femme est celle dont le visage figure sur cette photo, qui a été utilisée pendant pas loin de cinquante ans, pour diverses raisons, dont probablement l'inertie des scientifiques n'est pas la moindre…

Pour terminer cet éditorial, je me permet de vous rappeler que ce bulletin est le vôtre, il vous appartient, et il sera ce que vous en ferez. Alors, faites-nous profiter de votre dernière expérience pédagogique, de votre dernier algorithme, de votre dernière activité de recherche ; souhaitez-vous rendre hommage à une personne, rapporter un fait historique, donner votre positionnement sur un sujet d'actualité qui vous tient à cœur… ou bien tout simplement rejoignez le comité de rédaction de 1024…

Il ne me reste plus qu'à vous souhaiter une bonne lecture et beaucoup d'enthousiasme pour participer, vous aussi, à ce bulletin…

Denis Pallez
Rédacteur en chef
1024@societe-informatique-de-france.fr

Le mot du président

Fix the system, not the women… and fix the law?

Isabelle Collet[1], Yves Bertrand[2]

La professionnalisation de l'édition de 1024 grâce à nos partenaires de C&F éditions, dont je salue ici les valeurs qu'ils portent autant que le savoir-faire professionnel[3], est un événement important pour la SIF. Cela m'a incité à l'accompagner d'un « mot du président » portant sur un sujet de société central pour notre discipline : l'égalité femme-homme en informatique, tant elle demeure un objectif lointain.
Mais… aurais-je été, seul, fondé à discourir sur pareil thème ? Non ! Me vint alors l'idée de solliciter la référence en la matière, la chercheuse qui maîtrise pleinement le sujet dont elle a fait le centre de ses recherches depuis des décennies : Isabelle Collet. À deux claviers, donc, nous nous sommes autorisés à corédiger un « mot du président » au moins trois fois plus long que ceux que je concocte habituellement. Bonne lecture !

Fix the system, not the women est le titre du pamphlet de Laura Bates[4] qui montre de manière magistrale que cibler les femmes plutôt que le système qui engendre les inégalités de genre est une erreur majeure. En particulier les actions prétendant agir sur les femmes — *fix the women* — pour les inciter à investir le numérique n'ont jamais fait la preuve de leur efficacité[5]… au mieux.

1. Professeure en sciences de l'éducation de l'université de Genève, membre d'honneur de la SIF.
2. Université de Poitiers, président de la SIF.
3. « Typothérapie : fragments d'une amitié typographique », Nicolas Taffin, C&F éditions, janvier 2023.
4. « Fix the system not the women », Laura Bates, 2022, Simon & Schuster.
5. « Les oubliées du numérique », Isabelle Collet, éd. Le passeur, septembre 2019.

DOI:10.48556/SIF.1024.24.7

Au pire, elles sous-entendent que les femmes demeurent le problème alors qu'elles font partie de la solution ! Laura Bates pour plusieurs champs de la vie sociale, et Clémence Péronnet[6] pour les mathématiques et l'informatique prouvent que les seules actions efficaces sont celles agissant sur le système — *fix the system* — et non sur ses victimes.

Notre conviction ? Ces actions sont d'autant plus efficaces qu'elles sont portées par des avancées légales, voire constitutionnelles, qui pérennisent de nouvelles obligations d'égalité de genre. Elles relèvent d'un *fix the law*, qui encadre des *fix the system* pragmatiques. Nous étayons cette conviction en évoquant les dynamiques de deux pays en matière de place des femmes en informatique : l'Inde et la France. Deux pays qu'*a priori* tout éloigne… sauf un « détail » : dans chacun d'eux les femmes occupent environ 24 % des emplois du numérique. Pour l'Inde, ce taux s'inscrit dans une dynamique fortement positive fondée sur un tandem *fix the law — fix the system*, alors que pour la France, elle s'inscrit dans une dynamique gravement négative pourtant accompagnée de multiples *fix the women*. Explications.

L'Inde, le vélo, les filles et l'école

Moins de 25 €… c'est le prix d'un vélo dans le Bihar, État du nord-est de l'Inde. C'est le cadeau que fait le Bihar à ses écolières : depuis 2008, toutes ont un vélo, **offert** par leur État.[7]

Juste avant la mise en place de ce dispositif, 160 000 filles étaient scolarisées. Trois ans plus tard, leur nombre a plus que triplé pour atteindre 490 000 en 2011[8]. Plus l'origine sociale des filles est basse, plus l'effet de la mesure est fort. À la base de cette mesure, une volonté politique nationale forte et pérenne : réduire les inégalités de genre dans les états les plus pauvres de l'Inde. Notamment au Bihar sous l'impulsion de Nitish Kumar, ministre en chef depuis dix-neuf ans. Cette volonté s'ancre dans la constitution même de l'Inde et dans la loi *Right to Education Act* de 2009 qui rend l'école obligatoire et gratuite pour tous les élèves de 6 à 14 ans, et dont le premier effet fut d'augmenter significativement le taux de scolarisation, et en particulier celui des filles : plus de 96 % des élèves de 14 ans sont scolarisés en 2023[9].

À l'occasion de la parution d'une étude récente[10] sur les effets socio-économiques de la politique indienne nationale de mise à disposition de vélos aux enfants des zones rurales les plus pauvres, la presse française et internationale

6. « Matheuses », Clémence Peronnet, Claire Marc et Olga Paris-Romaskevich, CNRS éditions, jan. 2024.
7. « Pour les écolières indiennes, la "révolution silencieuse" est en route », *Courrier international,* 21 août 2024.
8. « How cycling set deprived Indian girls on a long-life journey », *The Guardian,* 25 nov. 2011.
9. « India's Annual Status of Education Report », 2023.
10. « A silent revolution: Rapid rise of cycling to school in rural India », Agrawal, Srishti & Seth, Adit & Goel, Rahul, juillet 2024, *Journal of Transport Geography,* Elsevier, vol. 119(C).

redécouvre en 2024 cette politique plongeant ses racines dans des grands programmes datant du début des années 2000. Deux à trois années du programme *Bicycle Distribution Scheme* (BDS) auront suffi à améliorer substantiellement la scolarisation des filles. En une génération, les programmes ayant donné naissance à la loi de 2009 et ceux qui la déclinent auront entraîné de tels progrès en matière de scolarisation que la presse les qualifie de « révolution silencieuse ».

En outre, il y a quelque chose de foncièrement pragmatique dans cette mesure. Ce vélo donne aux filles une autonomie et une confiance en elles : c'est mieux que de les conduire en voiture à l'école. Cela vaut peut-être même mieux que de construire une école au pied de chez elles. Ce vélo est une façon de dire aux filles qu'il est normal pour elles de sortir de chez elles, de se déplacer, de suivre des études loin de la maison. Et que leur pays a l'intention d'y mettre les moyens. Bien sûr, une partie de la mesure est de l'ordre du *fix the women*. Si le vélo protège les filles contre les agressions sexuelles dont elles peuvent faire l'objet sur le chemin de l'école, il ne fait rien contre les prédateurs qui considèrent que les écolières sont des proies faciles. Souvent les bonnes mesures sont hybrides, agissant à différents niveaux.

Cependant, avec l'insuffisance des moyens pour équiper les écoles dignement, les difficultés logistiques induites par la géographie montagneuse, le rôle des castes malgré la constitution de 1950 qui affirme l'égalité de tous les citoyens de l'Inde ces inégalités demeurent à la mesure du pays : immenses.

L'Inde, les femmes et le numérique

Et pour ce qui concerne les femmes et le numérique, l'Inde est une terre de contrastes. La part des femmes dans les formations ne cesse d'augmenter sous l'impulsion de politiques nationales volontaristes. En 2023, la part des diplômées dans les STEM (*science, technology, engineering, and mathematics*) est de 43 %[11]. Mais la part des femmes dans les emplois du secteur du numérique n'est que de 24 %. Très jeunes, les femmes se marient, ou plus exactement, sont mariées. Elles ont alors vocation, sous le joug des multiples biais de genre persistants, à rester à leur domicile pour s'occuper de leur famille. Ceci explique le décrochage entre le taux de femmes en formation et le taux de femmes en emploi dans le numérique.

Mais l'Inde n'a que fort peu vécu l'ère industrielle. Elle est passée directement à l'ère du numérique et est actuellement la troisième puissance numérique derrière les USA et la Chine. Pour les emplois, les biais de genre jouent à nouveau un rôle clé, mais en faveur du numérique cette fois. En effet, les métiers du numérique sont perçus comme ne nécessitant pas de force physique, comme peu qualifiés dans nombre de secteurs et… comme pouvant très

11. « Contrairement à la France, en Inde, les études de sciences attirent de plus en plus les filles », *AEF,* dépêche 689787, 3 avril 2024.

aisément être exercés au domicile pour aider les femmes à cumuler un emploi et la gestion de la famille !

La France, les femmes et le numérique

En France, la dynamique pour les femmes et le numérique agit en sens inverse de celle constatée en Inde. Depuis les années 80 la part des femmes ne cesse de baisser en informatique aux USA, en Europe et en France en particulier. En France, la part des filles dans les enseignements d'informatique dans le secondaire stagne à 15 %. Les effectifs de 2024 sont retombés au niveau de ceux de 2014.

La France a, elle, entièrement vécu l'ère industrielle. Aux côtés de la mécanique, de l'électronique, l'informatique a été perçue comme un nouveau secteur industriel et, comme les secteurs qui lui sont connexes, elle est toujours perçue comme une profession intrinsèquement masculine. Les biais de genre sont alors à l'œuvre dans les conseils que les parents, les prescripteurs en général, délivrent de manière très différenciée aux filles et aux garçons. Leur meilleur allié est l'absence de toute politique gouvernementale en faveur des sciences en général et de l'informatique en particulier depuis des décennies… il faut remonter à l'après-guerre et au premier septennat de De Gaulle[12] pour trouver une politique ambitieuse et cohérente pour la recherche et la science… y compris pour le numérique !

La part des femmes dans les formations en informatique du supérieur est aussi de 15 %. Les femmes occupent à peine 24 %[13] des emplois du numérique (32 % pour les STEM), alors qu'il y a quarante ans, elles en occupaient 35 %. Dans le secteur de la sécurité informatique, leur part a été divisée par presque deux sur la même période[14].

Jusque dans les années 70, le calcul et la programmation étaient perçus comme des activités peu valorisantes, mécaniques, répétitives… et donc confiés très largement aux femmes, qui se heurtaient à la fois à de solides plafonds de verre en matière de progression de carrière et à une invisibilisation massive de leurs contributions et de leurs succès. Mais dès que l'informatique acquit ses lettres de noblesses socialement et économiquement, les hommes investirent massivement le secteur. Et, sans quitter du tout le navire numérique, les femmes se sont retrouvées automatiquement en proportion de plus en plus ténue[15].

En amont de leur insertion professionnelle, le chemin de l'école des jeunes françaises est, au sens propre comme au figuré, incomparablement plus simple et sûr que celui des jeunes indiennes, au point qu'en France 56 % des lycéens

12. « Le Général de Gaulle et la recherche en France », *La Revue pour l'histoire du CNRS,* Pierre Lelong, 1999.
13. « Emploi, chômage, revenus du travail », *INSEE Références,* édition 2023.
14. « Les chiffres clés sur les femmes et la tech », *La grande école du numérique,* 2024.
15. « Non, les femmes n'ont pas déserté le numérique », Sara Bouchenak, *Blog Binaire du Monde,* 8 mars 2024.

sont des… lycéennes depuis des décennies. Mais c'est plus tard, pour celles qui deviennent scientifiques, que la situation se complique et freine d'autant leurs possibilités d'occuper des fonctions de cadre de haut niveau. En effet, l'enseignement supérieur est loin d'être la *safe place* qu'on aimerait qu'il soit. Sur le site de la fondation L'Oréal, on trouve une grande enquête sur ce qu'il se passe dans les laboratoires[16]. Les femmes racontent les remarques graveleuses, les mises à l'écart, les captations du travail et finalement, les agressions sexuelles. Il est évidemment beaucoup plus difficile d'être une excellente étudiante ou une scientifique de premier plan lorsqu'on doit gérer ce genre de situations en plus de la pression concurrentielle et de la course à l'excellence inhérente à l'Université. Le violentomètre, conçu par Giorgia Magni, présenté sur le site de L'Oréal et disponible aussi sur le site du G-RIRE de l'UNIGE[17] devrait être affiché dans tous les laboratoires…

Inde et France : des dynamiques inverses

La dynamique positive de l'Inde s'appuie sur deux éléments clés : une constitution, des lois (*fix the law*) ; et des programmes cohérents, ambitieux, déployés à long terme comme le programme BDS (*fix the system*) dont les actions sont simples (1 fille — 1 vélo), radicales (gratuité), systématiques (pour toutes les filles) . Notons qu'à aucun moment il n'est question de communication, d'actions de sensibilisation du public féminin, ni d'incitations en tous genres.

La dynamique négative de la France s'explique par deux éléments clés : nulle loi-cadre pour favoriser la formation et l'insertion professionnelle des femmes dans le domaine des STEM, et d'incessantes réformes précipitées de l'enseignement supérieur et secondaire, sans stratégie *a priori*, sans évaluation *a posteriori* avant d'engager la réforme suivante. Plusieurs centaines d'actions de communication incitent les femmes (*fix the women*) à choisir une carrière scientifique, et en particulier une carrière dans la « Tech ». Avec, sur près d'un demi-siècle, les succès que l'on sait.

Mais par rapport à l'Inde, la France — comme les autres pays du Nord — a une difficulté supplémentaire. En 2010, le Conseil fédéral suisse[18] s'est interrogé sur la pénurie de spécialistes en mathématiques, informatique, sciences naturelles et technique (MINT) Le rapport pointe une question Nord / Sud qui dépasse de loin la question de genre.

Une étude internationale s'appuyant sur les données de plusieurs dizaines de milliers d'élèves dans plus de 20 pays montre que les jeunes de 15 ans sont d'autant moins intéressés aux STEM que leur pays a un indice de développement

16. https://www.fondationloreal.com/fr/nos-programmes-pour-les-femmes-et-la-science/contre-les-violences-sexistes-et-sexuelles.

17. https://www.unige.ch/fapse/grire/outils-de-formation/violentometre.

18. « Pénurie de spécialistes MINT en Suisse. Ampleur et causes de la pénurie de personnel qualifié dans les domaines MINT », *Rapport du conseil fédéral,* Confédération suisse, 2010.

humain (IDH) élevé au sens des Nations Unies. On observe une corrélation négative linéaire entre l'indice de développement d'un pays et le souhait exprimé par les jeunes de 15 ans de devenir des spécialistes MINT. Dans des pays comme le Bangladesh, le Ghana ou l'Ouganda, où l'IDH est bas, les jeunes de 15 ans sont très favorables à des études MINT. Au Japon et en Europe occidentale, c'est exactement le contraire.

En effet, les pays du Sud ont une économie moins basée sur le secteur du service que les pays du Nord. Il y a donc plus de métiers possibles, notamment parmi ceux d'ordinaire très féminisés. Comparativement, les femmes dans les pays du Sud sont moins nombreuses dans les métiers du service et du soin aux personnes parce qu'il y a moins d'emplois rémunérés dans le service et le soin aux personnes. En revanche, dans les pays du Nord, le stéréotype qui inculque aux femmes que les mathématiques ne sont pas pour les filles reste très puissant. Dans les pays qui offrent une large palette de métiers du soin et des services en général, les femmes peuvent estimer stratégique de faire carrière dans ces métiers, peut-être moins rémunérateurs mais dans lesquels leur place ne sera pas continuellement contestée et dans lesquels elles arrivent plus facilement à se projeter.

Résumons. En Inde (et dans nombre de pays du Sud) : une dynamique positive, des politiques systémiques, une structure d'emploi favorable ; en France (comme dans nombre de Nord) : une dynamique négative, des politiques atones, une structure d'emploi défavorable.

Un projet national en France

C'est dans ce contexte qu'en France INRIA a été chargé de piloter le programme interministériel TechPourToutes[19]. Programme ambitieux s'il en est, avec des responsables totalement engagées, il vise à accompagner à l'horizon 2029 de l'ordre de 20000 étudiantes pendant leur formation supérieure dans la Tech. L'idée initiale — *fix the system* — était d'accompagner financièrement, à une hauteur très significative pendant cinq ans, 10000 jeunes femmes par an. Cette idée n'a pas franchi les fourches caudines gouvernementales : une discrimination positive consistant à financer fortement les études supérieures d'une population exclusivement féminine serait inconstitutionnelle ! Le projet a évolué en une fédération d'innombrables actions en faveur des femmes pour la Tech, avec pour but de renforcer l'efficacité de chacune d'elles. L'essentiel des actions concernées relève du *fix the women*.

Chacun de nous deux a alerté les responsables du projet TechPourToutes sur le danger à mettre l'accent sur le *fix the women* et sur l'intérêt à promouvoir des mesures systémiques, structurelles. Dans un établissement de formation secondaire par exemple, il s'agirait de sensibiliser et former l'encadrement

19. https://www.techpourtoutes.io/.

pédagogique et administratif, souvent majoritairement masculin plutôt que d'interpeler les filles et les enseignantes elles-mêmes.

Mais le *fix the law* existe en France, juste en amont du projet, dans l'enseignement secondaire. Il y est redoutablement efficace, négativement, et à court terme. En quatre ans, il a ramené la part des filles en sciences au lycée à un niveau inférieur à celui du début des années 90[20]. Le taux de filles en informatique a reculé d'un tiers. Les effectifs des filles en informatique sont revenus au niveau de 2014 : moins d'une fille par lycée étudie l'informatique en terminale en 2024[21]. Sans un autre *fix the law*, nul n'est besoin d'être grand clerc pour conjecturer que le taux et les effectifs des filles étudiant l'informatique au lycée seront étales pour longtemps, et que les projets *fix the women* les plus ambitieux et fédérateurs risquent d'engendrer des désillusions.

Fix the system : un exemple

Pourtant, le *fix the system* consistant à mettre en place des mesures pragmatiques, parfois modestes, mais systémiques s'avère aisément efficace. Évoquons-en un exemple porté par la SIF : le prix Babbage-Lovelace créé en 2023 en collaboration avec l'Académie des sciences. S'est immédiatement posée la question de l'attractivité d'un tel prix pour les femmes et par conséquent du taux de femmes lauréates au prix. Ce taux ne serait-il pas trop faible, au regard des tailles des viviers respectifs d'informaticiennes et d'informaticiens ? La question fut tranchée par un simplissime *fix the system* qui garantit une parité parfaite entre lauréates et lauréats. Il fut en effet décidé de décerner un prix « informaticienne » et un prix « informaticien » chaque année. De ponctuelles préventions à l'endroit d'une telle démarche affleurèrent : le prix « femmes » ne sera-t-il pas moins sélectif que le prix « hommes » ? Les femmes accepteront-elles de ne concourir qu'entre elles ? Même question pour les hommes, au demeurant. Ces atermoiements disparurent aussi vite qu'ils survinrent. Le nombre de lauréates est par construction même (*fix the system*) identique au nombre de lauréats. Si le nombre de candidates reste inférieur au nombre de candidats, leur taux fut dès la première année largement supérieur au taux de femmes en formation ou dans les métiers de l'informatique. Une fois encore, un renforcement systémique simple des opportunités offertes aux femmes agit bien plus sûrement et rapidement que toute campagne de communication !

20. « Filles et sciences : un problème majeur et persistant de la réforme du lycée », *Collectif Maths&Sciences,* 18 juin 2024.

21. « La spécialité "Numérique et sciences informatiques" au lycée en 2024 : péril en la demeure ! », *ce numéro de 1024.*

Fix the law : un impératif

Notre propos est donc le suivant. Arrêtons de dépenser temps et énergie pour des mesures de type *fix the women*, dont l'efficacité n'a jamais pu être mesurée. Promouvons des actions de type *fix the system*, structurelles, simples, massives, dont des pays aussi pauvres que l'Inde ont le secret, parce que pour eux, nécessité fait loi, au sens premier du terme : la nécessité absolue, perçue par certains dirigeants de sortir au plus vite leur pays de la misère se réifie en des lois ambitieuses.

Conférons donc aux actions systémiques un cadre législatif, voire constitutionnel, un *fix the law* à chaque fois qu'elles ont besoin, pour être efficaces, d'un horizon lointain et stable, de cohérence entre elles et d'un appui qui s'affranchit des contingences politiques, si mouvantes dans la France de 2024. Et surtout, assurons-nous que cette loi soit appliquée !

Portons par exemple à long terme l'idée que ***si l'informatique est enseignée au collège, elle doit l'être comme discipline à part entière,*** distinctement du programme de mathématiques et de celui de technologie, et obligatoirement enseignée par des professeurs d'informatique[22] <*fix the system 1*>.

Portons corollairement l'idée qu'***au lycée l'informatique pourrait faire partie du tronc commun en première ou terminale***[23] <*fix the system 2*>. Refusons toute idée d'un enseignement d'informatique plus « soft », moins technique au prétexte fallacieux d'intéresser les filles, en sachant qu'elles sont aussi douées que les garçons pour la technique, comme en témoigne l'histoire du calcul, des ordinateurs et de l'informatique du XX^e^ siècle.

Le MESR préconise à juste titre que tout diplôme de premier cycle inclue au moins 30 heures de cours pour 3 ECTS sur la transition écologique pour un développement durable[24]. Par analogie, portons en outre l'idée que ***tout diplôme de premier cycle pourrait intégrer au moins 30 heures de cours pour 3 ECTS sur la transition numérique*** pour un développement durable, responsable et de confiance[25] <*fix the system 3*>. Trois mesures systémiques simples, peu coûteuses, lisibles, ne nécessitant aucune communication puisque le public visé est la totalité des élèves et des étudiants, de façon obligatoire parce que l'acculturation au numérique, comme celle à l'écologie, devient un bien commun culturel des femmes et des hommes du XXI^e^ siècle.

Défendons enfin un indispensable *fix the law* qui imposerait l'***application de la loi « Rixain »***[26] ***dans nos établissements de formation,*** en particulier

22. « Inquiétude sur la formation en informatique des enseignants du collège », *communiqué de la SIF,* 17 août 2024.
23. « Informatique au lycée : péril en la demeure ! », *communiqué de la SIF,* 31 mai 2024.
24. « Former les étudiants de premier cycle à la transition écologique pour un développement soutenable », `https://www.enseignementsup-recherche.gouv.fr`.
25. « Pour un numérique durable, responsable et de confiance », CIGREF, `https://www.cigref.fr`.
26. Loi n°2021-1774 du 24 décembre 2021 visant à accélérer l'égalité économique et professionnelle.

dans le secteur de l'informatique. Comme les lois-cadres indiennes, cette loi se renforce à mesure que la société prend conscience de l'urgence à catalyser par de nouvelles lois[27] la longue marche vers l'égalité femme-homme. Leur déclinaison dans nos établissements de formation comme cadre des trois <*fix the system*> proposés serait plus efficace que des tombereaux de *fix the women* qui ne soignent que notre mauvaise conscience collective sans traiter les racines des problèmes.

Pour conclure ce trop long « mot du président », pour illustrer notre propos et provoquer le débat en l'appliquant à la SIF, nous vous jetons en pâture la proposition suivante : inscrivons dans les statuts de la SIF que son conseil d'administration et son conseil scientifique doivent être paritaires. Chiche ?

27. Loi n°2023-622 du 19 juillet 2023 visant à renforcer l'accès des femmes aux responsabilités dans la fonction publique.

Agrégation d'informatique 2023 Résultats et analyse

La SIF

Ce document résume le rapport du jury de la deuxième édition de l'agrégation externe d'informatique, qui s'est déroulée en 2023, sous la présidence de Sylvie Boldo[1]. La rédaction de ce rapport est très modulaire, ce qui a permis de rédiger le présent résumé en extrayant du document complet — avec l'accord des auteurs — des passages entiers très peu modifiés plutôt qu'en rédigeant un texte spécifique, qui aurait risqué de déformer le propos originel. Le rapport complet est disponible sur le site de l'agrégation[2]. La première section du présent résumé précise le déroulement du concours des statistiques sur les candidats, la deuxième section est dédiée aux épreuves écrites, et la troisième section aux les épreuves orales.

Déroulement et statistiques

Le rapport a plusieurs objectifs : c'est d'une part une analyse et un bilan de cette édition (statistiques sur les candidats, réussite aux épreuves, erreurs fréquentes, conseils…) et d'autre part un document à l'attention des futurs candidats et préparateurs (attentes du jury, écueils à éviter…).

Déroulement du concours

L'agrégation externe section informatique est régie par l'arrêté MENH2112-666A du 17 mai 2021, qui en définit le programme et les épreuves.

1. Directrice de recherche, Inria.
2. https://agreg-info.org.

DOI:10.48556/SIF.1024.24.17

Le programme des épreuves d'admissibilité se compose :
- des programmes d'enseignement de la spécialité « numérique et sciences informatiques » (NSI) de la première et de la terminale de la voie générale du lycée ;
- des programmes des classes préparatoires scientifiques aux grandes écoles « mathématiques, physique, ingénierie et informatique » (MP2I) et « mathématiques, physique, informatique » (MPI) ;
- d'un programme complémentaire élaboré par le jury.

Il a pour but de permettre au jury d'évaluer le recul des candidats sur les notions enseignées et est conçu par « agrégation » des programmes sus-cités. Pour tout le programme, il est attendu des candidats un recul correspondant au niveau master.

Les épreuves écrites se sont déroulées du 6 au 8 mars dans les académies, avec 5 heures pour la composition d'informatique, 6 heures pour l'étude d'un problème informatique et 6 heures pour l'épreuve spécifique. Les épreuves orales se sont déroulées du 17 au 22 juin 2023 au lycée Paul Valéry à Paris, organisées par le jury. Chaque admissible a été convoqué pour les trois épreuves orales sur trois jours consécutifs pour la leçon d'informatique (4 heures de préparation, 1 heure d'épreuve), les travaux pratiques de programmation (5 heures de préparation, 1 heure d'épreuve), la modélisation (4 heures de préparation, 1 heure d'épreuve).

Statistiques

Comme en 2022, le concours a été attractif, avec un très grand nombre de candidats. De plus, le niveau final des agrégés est excellent. Le jury a pourvu sans hésiter tous les postes, avec une liste complémentaire de 4 noms en plus des 22 places initiales, pour 48 admissibles. Les agrégés ont montré de très belles qualités, autant disciplinaires que pédagogiques. Le jury les a volontairement testés sur des domaines différents de l'informatique et des langages différents avec succès et a pu sélectionner des informaticiens complets, avec une vision large de la discipline.

Pour 2023, les épreuves écrites ont été communes avec l'agrégation d'informatique du Maroc (215 présents pour 511 inscrits au Maroc, et 174 présents pour 440 inscrits en France). Les candidats des deux concours ont passé les épreuves écrites en même temps. Leurs copies ont ensuite été corrigées par les mêmes correcteurs, anonymement et sans connaissance de la nationalité. Le chiffre des inscrits et des présents est en baisse par rapport à 2023, entre 20 et 25 %. L'agrégation d'informatique reste cependant l'un des concours les plus sélectifs de l'agrégation externe, avec environ 8,4 candidats pour un poste.

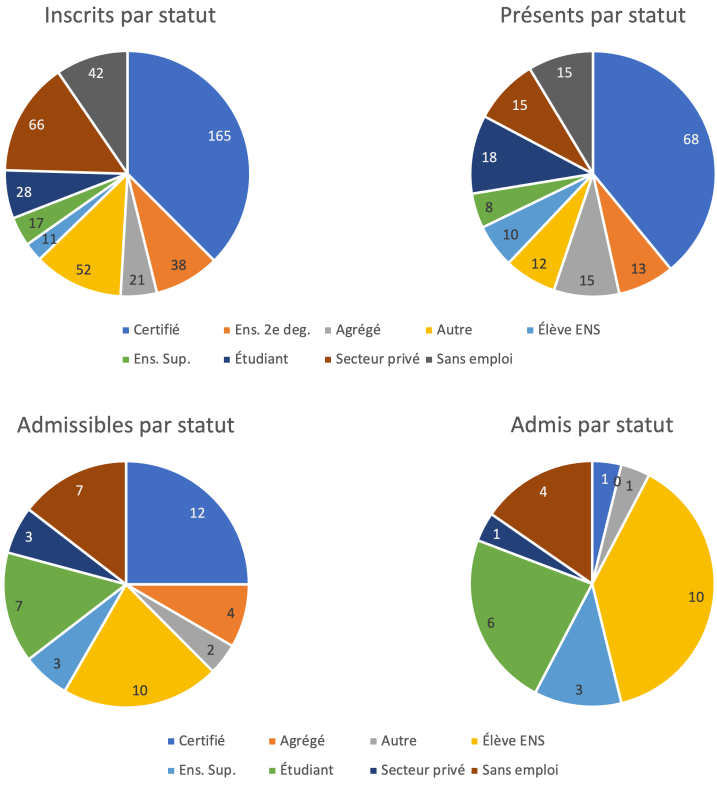
La moyenne des candidats présents aux trois épreuves écrites est de 4,68[3] et celle des candidats admissibles de 12,35, avec une barre d'admissibilité à 8,76.

3. Toutes les notes données sont sur 20 points.

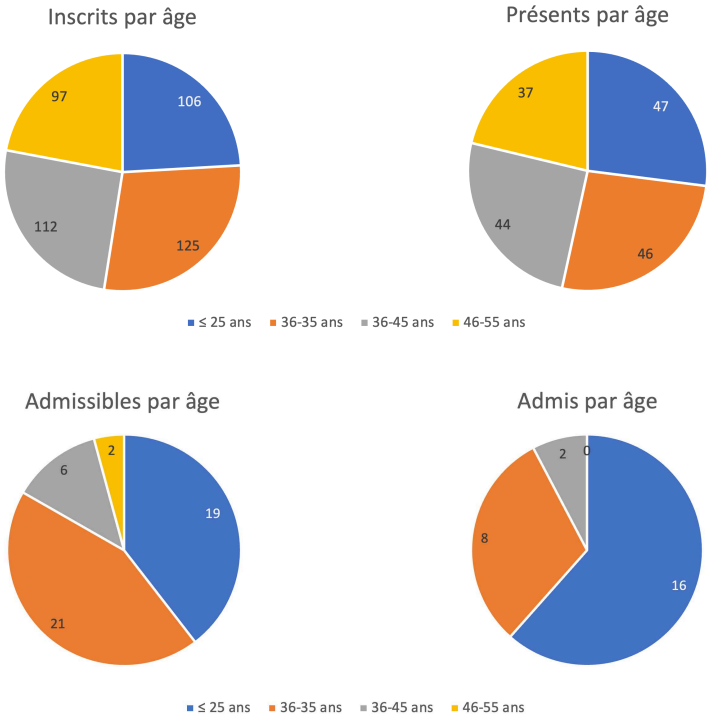
La moyenne des candidats présents aux trois épreuves orales est de 11,06, et la moyenne des candidats admis en liste principale de 13,32. Sur l'ensemble des épreuves, la barre d'admission est de 12,17 pour la liste principale et de 11,44 pour la liste complémentaire.

STATISTIQUES PAR STATUT

Les statuts sont donnés par les candidats de manière purement déclarative. Les profils des candidats sont extrêmement divers, même si cette diversité tend à se réduire tout au long du concours, la population des candidats bénéficiant de l'environnement d'une préparation devenant prépondérante à mesure que le concours se déroulait. Le profil des admis est un peu plus varié que l'an passé, même si les élèves des écoles normales supérieures constituent encore un bon tiers des lauréats.



STATISTIQUES PAR ÂGE



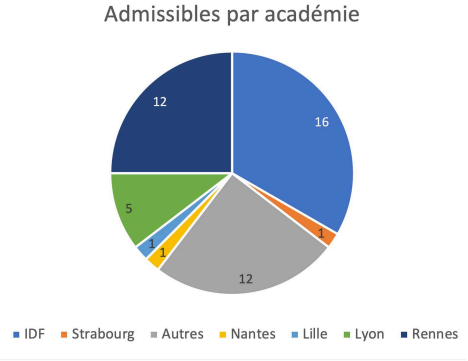
STATISTIQUES PAR GENRE

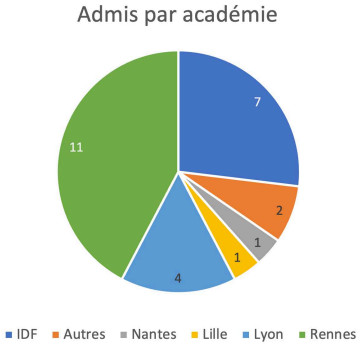
Le genre, déclaratif, ne permet à l'inscription que les deux choix M. ou Mme. La représentation des femmes en sciences et en informatique en particulier reste très faible. Le taux de femmes inscrites est faible (moins de 19 %), les femmes se présentent en moindre proportion, et elles ont plutôt moins bien réussi les épreuves écrites (étant entendu que les copies sont anonymes). Pour les épreuves orales, il est délicat de faire des analyses sur un échantillon aussi réduit, probablement corrélé avec d'autres facteurs statistiques mais ce point reste une préoccupation du jury, surtout au vu de la diminution du taux de femmes admises entre les sessions 2022 et 2023.

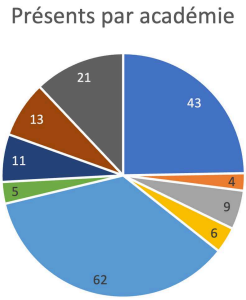
Genre	Candidats	Présents	Admissibles	Admis
M.	359	148	45	25
Mme	81	26	3	1

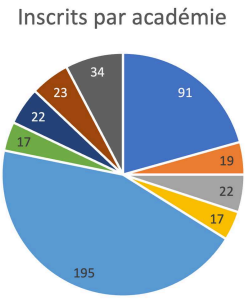
STATISTIQUES PAR ACADÉMIE

Les seules académies à proposer une préparation à l'agrégation externe d'informatique sont les académies de Créteil-Paris-Versailles, Lyon et Rennes, même si la première préparation peut se suivre partiellement à distance.









Épreuves écrites

L'architecture générale de l'écrit de l'agrégation externe d'informatique est définie par l'arrêté du 17 mai 2021. Dans les limites de la maquette définie par cet arrêté, le jury affirme sa volonté d'utiliser l'ensemble des épreuves à sa disposition pour recruter des informaticiennes complètes,[4] capables à la fois de réflexes sains dans un ensemble de domaines assez larges de l'informatique (épreuve 1), et dotées d'une solide maîtrise de la programmation et de l'algorithmique (épreuve 2). L'épreuve 3 est l'occasion de démontrer des connaissances approfondies dans un domaine au choix des candidates : l'ingénierie logicielle ou les fondements de l'informatique.

Certaines remarques s'appliquent à toutes les épreuves. En particulier, on attend d'une candidate aux fonctions de professeur qu'elle sache rédiger une copie. En conséquence, un passage rayé ou hachuré n'est jamais lu. De plus, quand plusieurs versions sont proposées par la candidate, la correctrice ne prend pas la meilleure. C'est à la candidate de choisir ce qu'elle présente à l'évaluation du jury.

Une préoccupation du jury est de valoriser les candidates qui sont précises sur les parties élémentaires du sujet plus que celles qui bâclent le début pour avancer plus vite vers les parties plus ambitieuses. Le jury recherche des candidates solides et précises sur les bases de la discipline, afin de pouvoir transmettre ces dernières. Il est attendu, pour l'écrit et pour l'oral, que les candidates maîtrisent le vocabulaire des mathématiques nécessaire à l'étude des éléments au programme de l'agrégation d'informatique, dont les algorithmes probabilistes et l'analyse de complexité.

Le jury tient compte dans la notation des épreuves de la maîtrise écrite et orale de la langue française (vocabulaire, grammaire, conjugaison, ponctuation, orthographe). Le jury attend donc une certaine tenue dans l'écriture, la présentation et l'orthographe. Ces attentes sont modérées, mais réelles. Le jury fait la part des choses entre ce qui est explicable par le contexte « concours » et ce qui risque d'handicaper une future agrégée dans sa pratique professionnelle et la transmission des savoirs. Les travers relevant de la seconde catégorie sont pris en compte, aux côtés de la maîtrise disciplinaire, dans l'évaluation d'une copie. Le jury a apprécié de lire des réponses construites, et rappelle qu'une phrase commence par une majuscule et se termine par un point, même quand elle est très courte et qu'elle constitue à elle seule une réponse à une question.

4. Pour alléger le texte, ce chapitre est uniquement écrit au féminin, mais cela doit être pris dans le sens du neutre : l'ensemble de ce texte, sans exception, concerne uniformément candidates et candidats.

Épreuve 1

185 candidates ; meilleure note : 18,09 ; moyenne : 6,51 ; écart-type : 3,89

Cette épreuve était composée de trois parties, balayant largement les bases de données, les graphes et l'architecture des ordinateurs.

Le jury a été déçu par les réponses à cette épreuve, car de nombreuses candidates ont peu abordé certaines parties. Hormis la sous-partie de la première partie, sur les requêtes SQL, les correctrices ont estimé que trop de questions ont été mal ou pas traitées dans cette épreuve. Pour cette épreuve en particulier, composée de plusieurs parties indépendantes, les correctrices souhaitent que les candidates notent le numéro complet de la question traitée et qu'elles ne mélangent pas les réponses aux différentes parties.

PARTIE I — PROVENANCE EN BASES DE DONNÉES

Cette partie a pour but de tester les connaissances en bases de données, notamment sur les requêtes SQL. Le sujet se poursuit par un point de vue plus logique permettant de parler de la provenance des résultats d'une requête.

Cette partie a été beaucoup traitée, les requêtes SQL étant souvent correctes. Le jury regrette la syntaxe SQL parfois approximative et le fait que vraiment peu de copies traitent la partie sur la provenance. Pour les requêtes SQL, les candidates doivent veiller à l'indentation pour aider à la lisibilité et à ne pas faire de jointure inutile. Il est aussi conseillé de bien lire les questions de façon à construire la requête demandée. nfin, certaines candidates ont utilisé des doubles négations dans les requêtes, rendant le code peu compréhensible et souvent faux. En ce qui concerne l'algèbre relationnelle — seulement au programme complémentaire —, le jury a toléré certaine souplesse dans la syntaxe utilisée.

PARTIE II — PONTS D'UN GRAPHE

L'épreuve portait sur l'implémentation, en langage OCaml, d'une structure de données permettant de déterminer efficacement les blocs d'un graphe non orienté (c'est-à-dire les composantes connexes restant après suppression des ponts, définis comme les arêtes dont la suppression fait croître le nombre de composantes connexes) dans le cas où le graphe est construit de façon incrémentale en ajoutant les arêtes une à une. Cette approche se base sur une variante de la structure classique de d'*Union-Find*(« Unir et trouver »), présente au programme de MPI.

Globalement, le jury regrette que cette partie de la première épreuve ait été si peu et si mal traitée, tant au niveau de la maîtrise du langage OCaml que du sujet lui-même, à savoir l'étude d'une structure de données. Ainsi, il est manifeste que les bases d'OCaml ne sont pas acquises pour une proportion non négligeable de candidates, alors que ce langage est explicitement au programme.

Dans certaines copies, on a pu lire du pseudo-code ressemblant à du Python, voire du code Python de manière totalement assumée. Cela n'est pas acceptable. Au début de l'énoncé, il était rappelé un ensemble de notations concernant en particulier la manipulation de tableaux. Il est donc regrettable de trouver dans des copies des notations incorrectes et une mauvaise utilisation des fonctions de base pour manipuler les tableaux.

Le jury a aussi eu l'impression que de trop nombreuses candidates pensent que programmer en OCaml revient à programmer exclusivement récursivement, et que toutes les fonctions doivent commencer par un *match … with*. Cette situation, naturelle lorsque l'on manipule des paramètres d'un type somme, ne correspondait pas à un grand nombre de questions de l'énoncé. D'ailleurs, un usage excessif et mal maîtrisé du filtrage peut être vu comme maladroit, voire créer d'inutiles complications.

Notons enfin que certaines questions attendaient de façon naturelle une solution itérative et que si l'utilisation de fonctions auxiliaires est admise voire recommandée, il est indispensable d'en expliciter clairement le rôle.

PARTIE III – ARCHITECTURE DES ORDINATEURS

Cette partie vise à évaluer les candidates sur l'architecture des ordinateurs, avec une première section sur l'algèbre de Boole, une deuxième sur les bascules et une troisième sur un automate de Moore.

La première section a été plutôt bien réussie et abordée par une majorité de candidates. La deuxième section concerne le fonctionnement des bascules RS et D et a été plus décevante. Les correctrices ont pu constater qu'une part importante des candidates n'a pas compris précisément le fonctionnement d'une bascule D. La troisième section, probablement la plus facile de la partie, a été complètement ratée. Il s'agissait uniquement de donner les tables d'un automate de Moore donné, pour ensuite le synthétiser à l'aide de portes NOR. La notion d'automate de Moore, au programme de l'agrégation, ne semble pas assimilée.

Épreuve 2

`179 candidates ; meilleure note : 20 ; moyenne : 5,80 ; écart-type : 4,76`

L'objet de cette épreuve est d'étudier la décidabilité du problème d'équivalence entre deux expressions rationnelles. Elle est composée d'une partie I qui contient des préliminaires algorithmiques, d'une partie II qui étudie le passage d'une expression rationnelle à un automate non déterministe équivalent, et d'une partie III qui porte sur l'équivalence de deux automates déterministes. Le but principal de l'épreuve est, en plus de vérifier les connaissances des candidates sur ces parties du programme de l'agrégation, de tester leurs capacités à raisonner sur des programmes informatiques.

Cette épreuve part de concepts et de résultats explicitement au programme de l'agrégation, pour aller jusqu'à des résultats classiques et des extensions tirées d'articles de recherche. Le parti pris concernant les résultats non explicitement au programme fut de les introduire et les analyser à travers le prisme de la programmation. Le sujet comportait régulièrement des questions de déroulement d'algorithmes sur des exemples simples qui devaient permettre aux candidates de se familiariser progressivement avec les notions introduites.

La partie I permettait aux candidates de prouver leur maîtrise de base de la programmation Python et de l'analyse de la complexité temporelle de fonctions simples, à travers l'étude de la complexité amortie pour les listes Python et le tri lexicographique.

La partie II commençait par une section visant à transformer une expression rationnelle en arbre. Il s'agissait ensuite d'appliquer l'algorithme de Berry-Sethi pour obtenir l'automate de Glushkov, et de prouver la correction totale du code fourni et sa complexité temporelle. Enfin l'énoncé s'inspirait de l'article « Passage d'une expression rationnelle à un automate fini non-déterministe » de D. Ziadi, J.-L. Ponty et J.-M. Champarnaud (Bull. Belg. Math. Soc. 4 (1997), p. 177-203) pour modifier la fonction de l'énoncé dans le but d'améliorer sa complexité. Cette partie permettait de tester les capacités des candidates à faire des raisonnements par récurrence et par induction.

La partie III portait sur la minimisation d'automates déterministes, pour tester l'équivalence de deux automates déterministes, en étudiant l'équivalence de Nérode, l'algorithme de minimisation de Moore, et en terminant par un algorithme de type d'*Union-Find*. Cette partie reposait sur les capacités des candidates à analyser des fonctions pour prouver des résultats sur les objets manipulés par ces fonctions. L'algorithme de type d'*Union-Find* est inspiré de l'article « A linear algorithm for testing equivalence of finite automata. » de J. E. Hopcroft and R. M. Karp (Technical Report 114, Cornell University, December 1971). Cette partie permettait aux candidates de montrer leur capacité à faire le lien entre les algorithmes et structures de données au programme et la théorie des automates.

PRÉSENTATION

Comme l'an passé, le jury invite les candidates à soigner la qualité scripturale, grammaticale et rédactionnelle de leur copie. Il y est doublement sensible : une copie convenablement écrite est plus facile à évaluer et, par ailleurs, il paraît inconcevable de recruter des enseignantes ne montrant pas une maîtrise raisonnable de la langue française. Il convient également d'utiliser un vocabulaire approprié : on ne dit pas d'une propriété qu'elle est « forcément » vraie, et il n'est pas davantage adéquat d'écrire « On est sur une complexité $O(n)$. ». De la même façon, il faut respecter la syntaxe et la graphie lorsqu'on écrit du code informatique : la casse n'est pas anodine en Python. Le jury n'apprécie pas le code

rempli de ratures ou le code sur deux colonnes. Il faut éviter les renvois d'une page à une autre avec des flèches, parce que cela pose des problèmes de propreté de la présentation. Enfin, il n'est pas utile de recopier la question : en rappeler le numéro suffit.

PROGRAMMATION

Le jury est surpris que certaines candidates ne sachent pas programmer en Python, langage pourtant au programme de l'agrégation. Dans les questions de programmation, le jury n'attend pas que le programme vérifie que les paramètres respectent les préconditions. Il n'est pas non plus utile de traduire du code en français.

FORMALISME ET RAISONNEMENT MATHÉMATIQUES

Comme l'an passé, le jury constate une très grande variabilité dans la qualité des raisonnements, qu'il s'agisse de leur présentation ou du fond. Certaines copies pourraient servir de modèle de corrigé. D'autres semblent rédigées par des candidates qui n'ont pas compris qu'elles passent un concours scientifique. Ainsi, un exemple ne peut servir à prouver la correction d'une fonction. Les symboles mathématiques doivent être employés à bon escient et ne se substituent pas à un verbe : on ne peut utiliser $\forall$ dans une phrase comme si c'était une abréviation. Enfin, il convient d'éviter de commencer une phrase par une formule, un mot-clé du langage de programmation ou un nombre.

Plus profondément, le jury déplore que certaines candidates emploient le symbole d'implication $\Longrightarrow$ en lieu et place d'une déduction. On ne peut, en tant que future enseignante, se permettre cet abus d'écriture qui peut aussi entraîner des erreurs de logique graves. Il convient d'employer le mot « donc » ou un synonyme pour affirmer chacune des propriétés en les déduisant les unes des autres. De même, la conclusion d'une réponse ne doit pas être introduite par une flèche, et moins encore par la flèche d'implication. L'absence de maîtrise de ce point fait désordre pour toute professeure scientifique et révèle un manque de recul sur les chapitres de logique du programme de l'agrégation.

La manipulation de la notation $O(\cdot)$ doit être raisonnablement rigoureuse. Une définition précise des notations O, Ω, Θ, doit être maîtrisée par les candidates à l'agrégation. Il est important que ces trois symboles soient bien différenciés les uns des autres dans l'écriture. Enfin, il convient d'être rigoureux lorsqu'on veut prouver une majoration de complexité par induction structurelle. On ne peut pas supposer que, pour une sous-expression E, la complexité est en $O(|E|)$. Cela n'a pas de sens. Il faut passer par une majoration avec des constantes explicites.

Épreuve 3

Cette épreuve permet à la candidate de choisir (à l'inscription) entre étude de cas informatique (107 candidates ; meilleure note : 17 ; moyenne : 7,30 ; écart-

type : 3,60) ou fondements de l'informatique (68 candidates ; meilleure note : 20 ; moyenne : 8,64 ; écart-type : 5,38).

Il y a eu d'excellentes copies dans les deux épreuves. Par contre, il apparaît que le vivier des candidates se présentant à chacune des épreuves n'est pas le même. Les deux épreuves ont donné lieu à des moyennes assez différentes. Cet écart se reflète dans les autres épreuves : les candidates ayant choisi l'étude de cas informatique ont eu une moyenne inférieure de plusieurs points à celle des candidates ayant choisi les fondements de l'informatique, et ceci sur les deux autres épreuves.

ÉTUDE DE CAS $(u_n)_{n\in N}$ ET $(\alpha_n)_{n\in N}$ INFORMATIQUE

Le sujet s'inspire d'un projet concernant des livraisons par des véhicules. La partie 1 présente la problématique et s'attache à des considérations algorithmiques et de programmation objet en Python. L'idée de la partie 2 était d'explorer un protocole applicatif de communication multisaut entre véhicules. Un grand nombre de réponses n'ont pas su faire la différence entre un protocole applicatif et le protocole TCP. La partie 3 explore la concurrence et la partie 4 concerne la gestion de bases de données. La partie 5 s'intéresse à la programmation Web.

Il est attendu que tout code intègre des commentaires faisant le lien avec le sujet, par exemple pour fusionnable et ses trois critères. On s'étonne que certaines candidates ne sachent pas écrire de code Python, JavaScript, HTML ou SQL, alors que tous ces langages sont au programme. Au-delà de la maîtrise syntaxique d'un langage de programmation, il est attendu que tout code soit lisible et intelligible. À cette fin, le recours à des fonctions et attributs, intermédiaires ou privés, avec un libellé explicite peut être une aide utile à la compréhension d'un code.

FONDEMENTS DE L'INFORMATIQUE

Le sujet de cette épreuve consistait à étudier différents modèles de réseaux de neurones formels, et à déterminer leur puissance en tant que modèles de calculs. Il visait à relier ce modèle à différents modèles plus classiques de l'informatique fondamentale : circuits booléens, automates finis, machines à compteurs, machines de Turing, et à voir que chacun de ces derniers modèles correspond à une certaine classe de réseaux de neurones formels.

Dans la partie I, on se focalisait sur le cas d'un unique neurone formel dont la fonction d'activation est la fonction de Heaviside (neurone à seuil). On devait dans un premier temps étudier comment calculer certaines fonctions booléennes, et montrer que des fonctions comme la fonction parité ne peuvent pas se représenter de cette façon. On discutait alors la représentation de ces fonctions, en utilisant des coefficients entiers. On étudiait ensuite la question de l'apprentissage des coefficients d'un neurone à seuil, en étudiant l'algorithme

classique du Perceptron proposé en 1957 par Frank Rosenblatt, en démontrant sa convergence selon les arguments proposés par Novikoff en 1962.

Dans la partie II, on commençait à étudier des réseaux de neurones, avec des questions de programmation en OCaml, et l'observation que ces réseaux permettent de coder les fonctions booléennes, et que l'on peut toujours supposer que les réseaux sont organisés en couches. Puis, on cherchait à établir qu'il y a des fonctions comme la fonction parité qui se calculent avec un réseau de neurones de façon beaucoup plus succincte qu'avec des circuits booléens.

Dans partie III, on montrait la NP-complétude du problème de l'apprentissage des coefficients d'un réseau de neurones à partir d'un ensemble d'exemples.

Dans la partie IV, on reliait les réseaux de neurones récurrents aux automates finis, en jouant avec différentes façons de donner les entrées à un calcul.

Dans la partie V, on reliait les réseaux de neurones dont la fonction d'activation est la fonction ReLU (unité linéaire rectifiée) aux machines à compteurs, et les réseaux de neurones dont la fonction d'activation est la fonction linéaire saturée aux machines de Turing.

La plupart des copies dénotent une très bonne compréhension du sujet et montrent un traitement convenable des parties I et II. La partie III y a été manifestement moins inspirante, en dehors des questions les plus élémentaires. La partie IV a été traitée par un nombre plus grand de candidates, cependant trop souvent sans rentrer suffisamment dans les détails. En particulier, les candidates ne distinguent pas avec suffisamment de clarté le raisonnement pour les différentes variantes possibles d'automates finis correspondants, alors que c'était parfois précisément l'objet de plusieurs des questions. Les parties IV et V ont été traitées globalement par très peu de candidates, en dehors des questions de programmation par les machines à compteur.

Globalement, trop de questions nécessitant du code Ocaml sont esquivées, ou traitées dans une syntaxe plus qu'approximative. La nature inclusive des bornes des boucles OCaml ne semble pas connue de nombreuses candidates. Trop d'inégalités strictes se transforment en inégalités larges au gré de la rédaction, et la manipulation des inégalités est trop souvent erratique, en particulier, en ce qui concerne les inversions d'inégalités lors de multiplications par des quantités négatives.

Épreuves orales

Le jury est conscient du stress et de la fatigue des candidats[5] avec des temps de préparation longs et des horaires parfois décalés. Son objectif n'est pas de piéger les candidats mais de les mettre dans un contexte leur permettant de montrer le meilleur d'eux-mêmes. Le jury a jugé l'oral de l'agrégation d'un très solide

5. Pour alléger le texte, ce chapitre est uniquement écrit au masculin, mais cela doit être pris dans le sens du neutre : l'ensemble de ce texte, sans exception, concerne uniformément candidates et candidats.

niveau global, à la fois en matière de connaissances disciplinaires et de compétences pédagogiques.

Le jury rappelle que les livres sont disponibles pour les préparations de toutes les épreuves. La liste des livres fournis par le jury est disponible sur le site de l'agrégation[6] et sera mise à jour pour la session 2024. Les préparations à l'agrégation ont elles aussi mis à disposition des livres, également accessibles à tous les candidats. L'environnement informatique mis à disposition des candidats est celui qui est téléchargeable sur le site du jury et les fichiers des candidats sont sauvegardés sur un serveur. Il est à noter que les responsables informatiques ne sont pas là pour aider les candidats à debugger leur code, mais peuvent intervenir en cas de problème avec l'environnement à disposition. Consulter la documentation disponible dans cet environnement doit être le premier réflexe.

De nombreux conseils et remarques s'appliquent à l'ensemble des épreuves orales et sont donnés dans le rapport. Les trois épreuves sont distinctes et permettent au jury d'évaluer des points différents. Il est donc requis des candidats qu'ils connaissent la description et les attendus des trois épreuves. Des temps maxima sont donnés pour les trois épreuves et rappelés au début de chaque oral. Il est important d'essayer de se rapprocher le plus possible de ces temps sans toutefois les dépasser : même s'ils sont indiqués comme des temps à ne pas dépasser, ils sont en fait des temps-cibles.

Le jury d'une épreuve d'oral n'a pas connaissance des résultats de l'écrit du candidat et s'impose une discipline stricte de ne pas discuter des prestations d'un candidat devant les autres membres du jury avant que celui-ci ait passé l'ensemble de ses épreuves. Lorsqu'un candidat se présente à une épreuve, il est ainsi assuré d'être examiné sans aucun préjugé. Il est à noter qu'en 2023, l'ensemble des interrogations étaient ouvertes aux visiteurs.

Leçon

45 présents ; meilleure note : 19,25 ; moyenne : 10,44 ; écart-type : 4,62

Le candidat se voit proposer (aléatoirement) deux sujets au choix ; il doit choisir de traiter l'un des deux, mais n'a pas à justifier son choix sur ce point. Il indique son choix au début de l'interrogation proprement dite et peut donc changer d'avis en cours de préparation (même si ce n'est guère conseillé). Le jury n'interroge pas sur le sujet qui n'a pas été choisi.

Si la leçon d'agrégation permet d'évaluer les compétences didactiques des candidats, elle est un oral scientifique avant tout. Cela a des conséquences sur les trois moments de l'oral : le plan doit comporter du contenu scientifique, et pas seulement l'annoncer ; le développement doit bien sûr montrer des qualités d'exposition, mais appliquées à un contenu scientifique clair et précis ; les questions posées par le jury peuvent porter sur la façon dont on expliquerait telle

6. https://agreg-info.org.

chose à tel type d'élève, mais elles portent également fréquemment sur le fond des notions abordées, au niveau de recul M2 attendu des candidats.

Sur chaque point, la maîtrise et le recul didactique sur le contenu scientifique sont appréciés et valorisés, mais il s'agit de la cerise sur le gâteau : il faut d'abord qu'il y ait un gâteau. Par ailleurs, l'agrégation est un concours de recrutement d'enseignants : une expertise sur des domaines hors programme ne se substitue pas à un recul pédagogique. Le document complet donne de nombreux conseils sur le plan et le développement de la leçon.

QUESTIONS

De nombreuses questions simples visent à vérifier que le candidat comprend et maîtrise ce qu'il écrit ; elles sont posées systématiquement et le candidat ne doit pas s'en trouver déstabilisé. Certaines questions peuvent porter sur l'enchaînement des éléments du plan ou sur sa cohérence interne. Le jury s'attend à des notations homogènes de la part du candidat, même s'il a utilisé plusieurs sources adoptant des conventions de notation ou des définitions légèrement différentes. Si le travail d'homogénéisation n'est pas fait dans le plan, le candidat sera interrogé et invité à le faire pendant l'échange avec le jury. Le candidat doit également faire attention au développement logique de son cours : il peut être maladroit d'utiliser des notions essentiellement introduites postérieurement dans le plan. Le jury ne manquera pas de questionner ces choix.

Les candidats doivent bien écouter les questions qui leur sont posées, et bien différencier, par exemple :

– comment expliqueriez-vous l'algorithme de Rabin-Karp à un bon élève de terminale NSI ? Ici, de nombreuses réponses sont possibles ; on peut envisager de dérouler l'algorithme à la main sur un exemple bien choisi ;
– pouvez-vous expliquer l'algorithme de Rabin-Karp Ici, on attend une réponse scientifique de même nature que pour un oral de master.

S'ils présentent une « méthode générale » dans leur plan, les candidats peuvent s'attendre à ce que le jury leur fournisse un exemple ou un cas d'usage à dérouler. Par exemple, dans une leçon sur la programmation dynamique où le candidat expose une manière générale d'obtenir une équation de récurrence entre sous-problèmes, le jury peut proposer sous forme d'un petit exercice un exemple de problème où il faut d'abord trouver cette équation et les bons sous-problèmes.

Travaux pratiques

`44 présents ; meilleure note : 20 ; moyenne : 12,38 ; écart-type : 4,33`

Cette épreuve orale a pour objectif principal l'évaluation de compétences pédagogiques des candidats.

L'épreuve se décompose en deux parties : les travaux pratiques proprement dits et la revue de code. Dans la première partie, le candidat doit produire et présenter du code, et justifier sa correction (à l'aide de tests, de preuves, etc.). Dans la seconde, le candidat est confronté à du code de mauvaise qualité et défectueux, comme pourrait en écrire un élève durant une séance de travaux pratiques, et doit proposer des corrections ou améliorations, comme le ferait un enseignant durant une telle séance.

Dans les deux cas, l'accent est mis sur la pédagogie et non sur la compétence disciplinaire ou la virtuosité, les codes à écrire ou à commenter étant conçus pour ne pas poser de difficultés techniques insurmontables aux candidats.

Comme pour les deux autres épreuves orales, on attend une conduite claire et organisée de l'exposé, avec mise en contexte (qui doit être présente sans être excessive) et présentation structurée des résultats. L'usage du tableau doit être au service de la compréhension du discours du candidat, et celui d'une vidéo-projection doit se concentrer autour de la projection de code, de l'exécution de tests et éventuellement de la présentation (éventuellement sous forme graphique) de résultats d'exécution. Une présentation basée sur des diapositives n'est pas souhaitée.

Le candidat doit veiller à respecter l'équilibre entre les deux parties de l'épreuve et, en particulier, ne pas passer trop de temps sur la revue de code. Le rapport fourmille de conseils pour la partie travaux pratiques et la partie revue de code.

Modélisation

45 présents ; meilleure note : 17,50 ; moyenne : 10,66 ; écart-type : 4,03

FORME ET STRUCTURE DE L'ORAL

L'oral se décompose en deux parties. La première partie consiste en un exposé de 35 minutes, qui prend la forme d'un cours d'ouverture à destination d'élèves de niveau L1, L2 ou CPGE. Le cadre d'énonciation repose sur une triple fiction : le jury n'a pas connaissance du texte proposé au candidat ; le candidat a trouvé ce texte tout seul lors de recherches pour préparer son exposé ; le candidat a formulé lui-même les pistes de réflexion qui sont proposées à la fin du texte.

On ne demande pas un exposé sur le texte, mais sur le sujet traité par le texte. Il n'y a donc pas lieu de citer le texte et encore moins de s'y référer. L'exposé doit être structuré et il est fortement conseillé de présenter son plan en début d'oral après une courte introduction. L'exposé doit aussi inclure la présentation argumentée d'illustrations sur ordinateur, et la discussion d'une dimension éthique, sociétale, environnementale, économique ou juridique telle que mentionnée par l'arrêté. La deuxième partie de l'oral consiste en un échange avec le jury.

CONTENU DE L'EXPOSÉ

Il est attendu des candidats d'une part une présentation et une discussion du problème et de sa formalisation, d'autre part une capacité à développer, compléter, voire améliorer les ébauches de solution esquissées. Par exemple, lorsque l'on fait des études de complexité asymptotique d'algorithmes présentés pour être appliqués à des cas concrets, le jury attend que le candidat soit capable de raisonner en ordres de grandeur.

Outre la capacité à mobiliser ses connaissances dans un contexte concret, le jury cherche également à évaluer la communication scientifique et pédagogique, c'est-à-dire la capacité de tenir un discours structuré, cohérent et pédagogique sur un sujet non trivial. Pour cela, le candidat doit construire un exposé fondé sur le texte fourni. Il ne s'agit pas de faire une lecture commentée du texte, mais bien de construire un exposé autonome. En particulier, les candidats ne doivent supposer du jury aucune connaissance préalable du texte. Le jury n'attend pas un traitement exhaustif de tout ce qu'abordent le texte et les pistes, mais l'esprit du document doit être respecté.

L'évaluation de l'utilisation du tableau est partie intégrante de l'évaluation des compétences didactiques et pédagogiques des candidats. Le jury attend que le candidat montre sa capacité à utiliser le tableau pour transmettre des notions et expliquer des exemples bien choisis.

Les illustrations doivent s'intégrer à l'exposé de manière cohérente et fluide. L'exercice porte sur la communication ; on attend donc une présentation efficace d'illustrations qui éclairent un aspect du texte. Le jury est plus sensible à la pertinence de l'illustration qu'à la complexité de l'implémentation, même s'il reste attentif à la qualité du code présenté. Le candidat peut insister plus ou moins sur la programmation mais le reste de l'exposé doit rester équilibré et contenir une présentation détaillée et précise d'une solution esquissée par le texte. En cas de programmation, le jury attend que le candidat exécute ses programmes pour vérifier leur fonctionnalité.

Le jury est attaché à la capacité des candidats à prendre du recul et envisager l'informatique comme une discipline et non comme un patchwork de domaines ; la capacité à s'appuyer sur le texte pour faire apparaître des liens dans d'autres sous-domaines que celui étudié par le texte constitue une vraie preuve de maturité scientifique. Le discours doit toutefois rester dans le cadre du texte, et ce dernier ne doit pas servir de prétexte à une digression dans un domaine où le candidat se sent à l'aise.

Les candidats doivent intégrer une discussion synthétique d'une dimension éthique, sociétale, environnementale, économique ou juridique. Cette ouverture ne suit pas nécessairement une piste proposée par le jury, toute initiative personnelle pertinente ayant d'ailleurs vocation à être valorisée.

L'épreuve de modélisation nécessite une réflexion préalable des candidats sur ses différents aspects : construction d'un exposé, utilisation du tableau,

illustration informatique, conclusion du propos (voir les exemples de sujets donnés aux sessions 2022 et 2023 publiés sur le site de l'agrégation[7]).

QUESTIONS

L'interrogation se poursuit par les questions du jury, qui peuvent porter sur toute la présentation du candidat, sur tous les sujets du programme se rattachant à celle-ci, ou sur les dimensions éthiques, sociétales, environnementales, économiques ou juridiques en lien avec le texte. C'est l'occasion pour le jury d'affiner sa perception de la compréhension du texte par les candidats, de faire préciser ou corriger des points, de fournir des indications pour que le candidat développe certains aspects qui lui ont résisté, ou encore de proposer des pistes de réflexion, des prolongements, ou des variantes du problème étudié.

Comme pour les séances de questions des autres épreuves, il est primordial que le candidat adopte une posture facilitant les échanges. Il doit écouter les questions jusqu'au bout et répondre de façon concise et la plus précise possible. Bien que cela soit parfois nécessaire, il est difficile et désagréable pour le jury de devoir couper la parole d'un candidat qui s'est lancé dans une réponse très longue.

REMARQUES GÉNÉRALES

L'épreuve de modélisation de cette session reposait sur deux sujets portant sur des thèmes différents : l'analyse de complexité en prenant en compte la prédiction de branchement et l'impact de la réduction du nombre de couleurs sur la taille des images stockées au format *png*. Les candidats ont plutôt bien compris le format de l'épreuve. Il reste tout de même quelques exceptions qui poussent le jury à rappeler des conseils d'ordre général (le rapport contient des remarques spécifiques à chaque sujet).

- Il est important de présenter un plan de son exposé dans son introduction.
- L'épreuve de modélisation n'est pas une seconde épreuve de TP, les illustrations informatiques ne doivent pas représenter l'essentiel de la présentation. En revanche, elles doivent servir le propos, mettre en lumière des phénomènes et ne pas être considérées comme de simples exercices de programmation.
- L'épreuve de modélisation est une épreuve d'informatique, même si les interactions avec d'autres disciplines scientifiques sont très intéressantes, ce sont bien les dimensions informatiques qui doivent être au cœur de l'exposé. En particulier, il n'est pas judicieux de passer trop de temps sur les aspects très mathématiques de certains sujets.
- La gestion du tableau est un enjeu important de l'épreuve. Il ne faut pas utiliser le tableau comme une feuille de brouillon, mais bien comme un support pédagogique.

7. https://agreg-info.org.

– La gestion du temps est également un enjeu important de l'épreuve. Il est par conséquent déconseillé de passer beaucoup de temps à recopier du code fourni dans le texte.

Le jury a été satisfait du niveau global des candidats. Malgré tout, certaines maladresses et lacunes sont apparues à plusieurs reprises.

– Le jury attend une posture d'enseignant de la part des candidats. Ils doivent en particulier parler et écrire dans un français correct, en utilisant un vocabulaire précis et rigoureux. Par exemple, les termes « aléatoire » et « optimal » ont été utilisés de façon très maladroite. En particulier, l'optimalité n'est pas une échelle de valeur. On n'est pas plus ou moins optimal.
– Certains candidats ont montré des fragilités concernant l'analyse de complexité et les ordres de grandeur. Le logarithme en base 2 et la racine carrée ne doivent pas être confondus.

Bacheliers : le Parcoursup du combattant

La SIF

Ce document analyse la gestion des candidatures des bacheliers aux formations du supérieur sur la période non informatisée d'avant 2009, sur celle d'« Admission post-bac » (APB, 2009-2017), et sur la période actuelle « Parcoursup » (depuis 2018). Il montre qu'un défaut de maîtrise par les pouvoirs publics du fonctionnement technique et des fondements scientifiques de dispositifs — comme Parcoursup — dont ils assurent la maîtrise d'ouvrage conduit ces dispositifs et les politiques publiques qu'ils incarnent à l'échec. Il propose une mesure pragmatique pour éviter que Parcoursup ne subisse l'abandon prématuré, coûteux humainement et financièrement de feu son prédécesseur : la hiérarchisation officielle, à la carte, des vœux des candidats pour accélérer significativement le fonctionnement du processus. Parcoursup est en effet un outil à qui il manque peu pour être accepté sur le long terme : un lien permanent entre scientifiques et politiques qui permette d'y introduire continûment des correctifs au bénéfice exclusif de tous ses usagers, et de moins subir les foudres populistes, des politiciens de toutes obédiences, plus actuelles et radicales que jamais, pour s'attirer les faveurs des jeunes électeurs.

Jusqu'à 2008 : du papier, des enveloppes et des timbres

Avant 2009, les candidatures des bacheliers à des formations du supérieur s'effectuent avec des dossiers papier et des envois postaux[1]. La procédure est

1. https://interstices.info/dapb-a-parcoursup-quelles-methodes-daffectation-post-bac/.

DOI:10.48556/SIF.1024.24.35

chronophage et fastidieuse. Un dossier de candidat passe de formation en formation, par rangs croissants des vœux des candidats. Chaque formation a connaissance du rang auquel elle se situe dans les vœux d'un candidat, et elle peut tenir compte de cette connaissance pour adapter sa réponse au candidat. Et notamment lui répondre négativement si elle considère que le candidat lui a attribué un rang trop bas. Réciproquement, les candidats peuvent ordonner leurs vœux plus stratégiquement que sincèrement pour influencer les réponses des formations. Un tel système est dit *manipulable*, parce qu'un comportement stratégique peut s'avérer plus efficace qu'un comportement sincère.

De 2009 à 2017. Admission post-bac, de vie à trépas

La mise en place d'APB avait pour but de simplifier, d'accélérer et de rendre plus juste l'appariement des bacheliers et des formations. APB est constitué d'un noyau : l'algorithme de Gale-Shapley (AGS), dit des *mariages stables*, utilisé pour associer candidats et formations (voir encadré ci-dessous décrivant l'algorithme). À condition de faire jouer aux candidats le rôle des prétendants et aux formations celui des courtisées, l'AGS semble paré de toutes les vertus : a) il est simple et rapide, b) il construit des appariements qui évitent les injustices[2], c) il privilégie les préférences des candidats par rapport à celles des formations ; d) il favorise les ordonnancements de vœux sincères.

APB est aussi constitué d'une procédure qui classe les candidats d'une formation selon trois principes :

– il intègre une proportion définie de candidats extérieurs à l'académie de la formation considérée ;
– il prend en compte, pour classer un candidat, le rang de son vœu pour cette formation ;
– il recourt au tirage au sort pour des formations non sélectives mais en tension.

Le premier principe est un choix politique qu'il ne nous appartient pas de discuter. Le deuxième principe dénote une incompréhension du fonctionnement même de l'AGS de la part des pouvoirs publics. En effet, le classement d'un candidat par une formation ne doit dépendre en aucun cas du rang de cette formation dans les vœux du candidat pour que l'algorithme fonctionne, parce que cette dépendance rompt la stabilité des appariements effectués. L'existence du principe fut une erreur politique majeure en trois actes :

2. La « stabilité » des couples créés se traduit par le fait que si une pour une formation F qui n'a pas retenu un candidat C, il n'existe jamais de situation telle que F aurait préféré C à un autre candidat et réciproquement C aurait préféré F à la formation qu'il a obtenue. Soit la formation, soit le candidat n'ont alors pas intérêt à tenter de modifier l'appariement proposé par l'AGS : les appariements proposés sont donc stables.

– acte 1 : la loi de 1984[3] entre en vigueur à un moment où les formations du supérieur absorbent sans difficulté tous les bacheliers (moins de 250 000) qui souhaitent poursuivre leurs études. Elle ne prévoit ni n'intègre la massification de l'enseignement supérieur ;
– acte 2 : APB (environ 537 000 bacheliers en 2009) prévoit techniquement un tirage au sort pour une formation non sélective qui ne pourrait accueillir tous ses candidats. Mais les pouvoirs publics ne prévoient pas ce cas politiquement. Tant que le tirage au sort ne concerne qu'un nombre non significatif de candidats, le tirage au sort opère à bas bruit ;
– acte 3 : en 2017, la massification de l'enseignement supérieur fait son œuvre (plus de 100 000 bacheliers de plus qu'à la naissance d'APB) : soudainement, le nombre de candidats concernés par le tirage au sort pour une seule formation peut se compter par centaines. Les injustices afférentes éclatent au grand jour.

L'ALGORITHME DE GALE-SHAPLEY

Appliqué à n hommes (prétendants) et n femmes (courtisées) ⋆ qui souhaiteraient se marier, l'algorithme suppose que toute femme exprime des préférences ordonnées sur tous les hommes et que ces derniers font de même sur toutes les femmes. Ses inventeurs, David Gale et Lloyd Shapley, ont montré en 1962 qu'il possède plusieurs propriétés : a) il se termine toujours en un temps raisonnable pour un ordinateur ⋆⋆ ; b) il trouve toujours une solution qui associe de façon dite *stable* ⋆⋆⋆ chaque couple créé ; c) il est optimal-positif pour les hommes : chaque homme est associé à la femme de meilleur rang possible dans ses préférences en préservant la stabilité de tous les couples ; d) il est non manipulable : un comportement sincère est plus efficace que tout comportement stratégique. Mais cet algorithme est dissymétrique. Il est en effet optimal-négatif pour les femmes : chaque femme est associée à l'homme du moins bon rang possible en préservant la stabilité de tous les couples ⋆⋆⋆.

⋆ Bien sûr, il est possible d'inverser les rôles joués par les femmes et les hommes ! ☺

⋆⋆ L'algorithme effectue n^2 itérations : il ne faut que quelques minutes à un ordinateur pour l'exécuter même si n est très grand.

⋆⋆⋆ Il n'existe aucune paire d'individus (x, y) non mariés entre eux tels que x préfèrerait être marié à y et y à x.

Ce dernier acte scelle le sort d'APB. Techniquement, pour l'AGS, le tirage au sort est la négation même d'un ordonnancement des candidatures. Sociétalement, c'est un parangon d'injustice. Les mécontentements induits sont renforcés par le refus du MESR d'être transparent sur l'algorithme et le code source

3. Loi n°84-52 du 26 janvier 1984 sur l'enseignement supérieur, article 14.

d'APB. Ils précipitent en quelques mois en 2017 l'abandon d'APB, pourtant fondé sur l'AGS dont les auteurs ont montré dès 1962 qu'il possède toutes les propriétés pour réaliser des appariements satisfaisants. Que révèlent ces manquements ? Que lorsque le politique méconnaît le fonctionnement d'un dispositif technique dont il assure la maîtrise d'ouvrage, qu'il n'anticipe pas l'obsolescence d'une loi ancienne, et qu'il s'enferre dans une réaction non adaptée à la situation, il voue la mise en œuvre du dispositif à un échec assuré, indépendamment de ses qualités intrinsèques.

APB post-mortem

Le 15 février 2018, l'OPECST[4] remet à l'Assemblée nationale et au Sénat son rapport[5] : *Les algorithmes au service de l'action publique : le cas du portail APB*. Tout y figure : l'analyse politique et technique des difficultés de mise en œuvre d'APB ; l'esquisse du fonctionnement de son éventuel successeur ; l'anticipation des biais politiques et techniques de sa mise en œuvre. Les débats sont détaillés, argumentés, rationnels. Ils émanent de politiques très au fait des conséquences sociétales de la mise en œuvre d'APB et de plusieurs scientifiques qui font autorité en matière de mathématiques, d'informatique, d'algorithmes d'appariements, d'économie. Sans l'expliciter, ce rapport constitue l'oraison funèbre d'APB. Dès le 15 janvier 2018, son successeur, Parcoursup, est déployé, avant même la promulgation de la loi ORE[6] qui ne le légalise que deux mois plus tard.

Depuis 2018 : le règne de Parcoursup

Parcoursup est constitué du même noyau national qu'APB : l'AGS, qui a fait ses preuves, au sens mathématique du terme, dès sa conception. Les pouvoirs publics conservent donc ce noyau à juste titre. Ils remplacent le classement concernant une formation, automatique, sans critère académique et avec tirage au sort, par un classement par une commission d'enseignants, sur critères académiques, sans tirage au sort, en l'asservissant à un taux minimal de candidats boursiers et à un taux maximal de candidats hors-académie.

Affirmons-le, quitte à passer pour des supporteurs du MESR : confier la responsabilité du classement des candidats à un groupe d'enseignants de l'établissement dispensant la formation souhaitée, et fonder ce classement sur des critères académiques nous paraît être la bonne, sinon l'unique façon de procéder. Elle est en outre techniquement pertinente, redonnant à l'AGS l'ordre total nécessaire à son fonctionnement sur les candidats à la formation considérée. Une autre manière de procéder existe, qui consiste à donner à toute formation non sélective les moyens d'accueillir tous ses candidats. Cette solution aux airs

4. Office parlementaire d'évaluation des choix scientifiques et techniques.
5. Compte-rendu de l'audition publique du 16 novembre 2017.
6. Loi du 8 mars 2018 relative à l'orientation et à la réussite des étudiants.

de vœu pieux dépasse le cadre de notre propos. Enfin, le MESR a eu raison de s'appuyer sur l'expertise incontestée de Claire Mathieu, directrice de recherche au CNRS à l'IRIF[7] et d'Hugo Gimbert, chargé de recherche au CNRS au LaBRI[8] De faire de la transparence de l'algorithme de Parcoursup et du code associé une exigence ne souffrant débat. De concevoir Parcoursup comme une déclinaison rigoureuse et vérifiable des attendus de la loi ORE. D'associer Hugo Gimbert et Claire Mathieu à la réflexion sur la finalisation de cette loi pour parfaire cette déclinaison.

Parcoursup contesté

Malgré ses atouts Parcoursup n'échappe pas à la critique. À chaque échéance électorale, simplisme et populisme conduisent les candidats de tous bords à réclamer sa tête. Les critiques d'opacité opposées à APB rejaillissent sur les travaux des commissions d'examen des vœux, au point que le Comité éthique et scientifique de Parcoursup (CESP) y consacre ses deux premières recommandations en 2023[9]. En effet, lorsque des enseignants de l'enseignement supérieur débordés œuvrent sous le joug de délais courts pour classer des milliers de candidats, que croyez-vous qu'il advienne ? Ils recourent, et c'est normal, à une procédure de classement automatisée des candidats. Les critères utilisés sont numériques : ce sont le plus souvent des moyennes pondérées des notes obtenues dans les disciplines en adéquation avec la formation visée. Les éléments qualitatifs comme les lettres de motivation sont-ils pris en compte ? Parfois. Et le lycée d'origine ? Peut-être. Les traces laissées par le refus du MESR d'être transparent sur APB sont si profondes et la notion d'algorithme, si peu comprise, que la vindicte politique et citoyenne est prête à fondre sur les jurys que constituent les commissions de classement. Leur souveraineté, leur transparence, leur impartialité sont contestées dès qu'elles usent de procédures automatisées, aussi pertinentes soient-elles. Une autre fragilité du système apparaît avec la massification de l'enseignement supérieur et la multiplicité des candidatures. Les candidats à une formation sont si nombreux que la différence entre le dernier accepté et le premier refusé est numériquement infime et donc non significative d'un quelconque écart de mérite académique. Et la situation est pire encore quand l'écart est nul : l'ordre total du classement est perdu.

Les vœux des candidats : non classés !

Revenons sur une évolution importante de Parcoursup par rapport à APB : les candidats ne font plus que 10 vœux non ordonnés alors qu'ils en effectuaient jusqu'à 24, tous ordonnés, avec APB. Le deuxième principe (cf. page 1) d'APB est

7. Institut de Recherche en Informatique Fondamentale, université Paris Cité — CNRS.
8. Laboratoire de recherche en informatique, université de Bordeaux — CNRS.
9. Comité éthique et scientifique de Parcoursup, rapport de février 2023.

à juste titre éliminé par Parcoursup : les formations ne tiennent plus compte des préférences des candidats… puisque ces préférences n'existent plus. Les pouvoirs publics intègrent donc le fait que le classement des candidats par une formation ne doit jamais tenir compte des préférences des candidats pour les formations. Mais ils traduisent cet impératif en supprimant toute possibilité, pour les candidats eux-mêmes, d'ordonner leurs vœux ! En oubliant le fait que le cœur de Parcoursup ne fonctionne qu'avec des vœux totalement ordonnés, et qu'il suffit de les masquer aux commissions locales pour éviter qu'elles tiennent compte de cet ordre. Une autre hypothèse existe. En effet, les pouvoirs publics sont sommés[10] par la CNIL en 2017 de ne pas baser l'affectation des candidats uniquement sur un algorithme et d'y introduire une intervention humaine. C'est chose faite avec les classements effectués par des enseignants. Mais quoi de mieux, en apparence, pour remettre de l'humain-usager au cœur du système d'appariement, que de faire interagir ce système avec les candidats eux-mêmes pendant une partie de l'été ? Rien, au vu du fonctionnement actuel de Parcoursup. Explications.

Parcoursup : mon manège à moi, c'est toi !

Les candidats indiquent donc leurs vœux à Parcoursup sans ordre. Comme Parcoursup dispose des classements complets des formations, les bons candidats obtiennent des réponses positives rapidement. Mais, sans disposer des préférences des candidats, Parcoursup s'arrête. Les premiers destinataires de réponses positives lui répondent en choisissant une formation parmi celles qui leur ont répondu positivement. Ou attendent un peu. Ou changent d'avis. Ou les deux simultanément. Fort de ces réponses, Parcoursup effectue un nouveau tour de carrousel. Et ainsi de suite, « au fur et à mesure et en continu », comme ne craint pas de le mentionner le calendrier officiel de Parcoursup de 2024. Parcoursup s'anime, tel un gigantesque carrousel effectuant de nombreux tours du 30 mai au 12 juillet 2024, avec les millions de réponses des candidats comme usagers captifs. Avec la phase d'admission complémentaire, ouverte dès le 11 juin et achevée le 12 septembre — pendant que nombre d'étudiants auront déjà entamé la deuxième semaine de leurs études supérieures ! — Parcoursup baguenaudera tout l'été.

Parcoursup s'améliore chaque année…

Il convient cependant de souligner que Parcoursup progresse chaque année. En 2018, il a mis 108 jours à rendre la totalité de ses verdicts, et seulement 44 jours en 2022 puis 37 jours en 2023. Début juillet, en attendant les résultats du

10. Décision MED-2017-053 du 30 août 2017, l'État mis en demeure par la Commission Nationale Informatique et Libertés.

baccalauréat, Parcoursup demandera aux candidats ayant des vœux en attente de… les classer par ordre de préférence. Le but est à juste titre d'accélérer la phase d'admission. Chaque année, à pas de velours, les possibilités partielles d'ordonnancement sont offertes à des périodes de plus en plus précoces : entre le 15 et le 18 juillet en 2022, entre le 30 juin et le 3 juillet en 2023. Le CESP dans son rapport de 2024[11] propose, voir encadré ci-dessous, d'avancer encore ces dates pour accélérer encore la procédure. Le calendrier 2024 de Parcoursup choisit de n'en pas tenir compte et retient la même période qu'en 2023.

… mais l'été des candidats demeure une gabegie sociale

Pour les moins chanceux, la phase complémentaire durera donc encore tout l'été, à subir du stress pour offrir de l'interaction humaine à Parcoursup. Ce carrousel apparaît comme une gabegie sociale dont l'objet est de le faire fonctionner, pas à pas, pour donner l'illusion d'un contrôle humain, alors qu'il est fondé sur un algorithme éprouvé depuis des décennies dans plusieurs pays. Cette illusion est d'autant plus douloureuse — si l'on en croit les statistiques — que les candidats de catégories socio-professionnelles (CSP) favorisées reçoivent des réponses positives très tôt durant la phase principale, et que ce sont les CSP défavorisées qui font le plus les frais de cette attente estivale.

11. Comité éthique et scientifique de Parcoursup, rapport de mars 2024.

LE CESP S'EXPRIME À PROPOS DE LA HIÉRARCHISATION

« [Il existe] un effet significatif dans les trois jours suivant la hiérarchisation en 2023 par rapport à 2021 et 2022 où il n'y avait pas de hiérarchisation à cette date. Le 3 juillet 2023, 11 500 candidats admis, mais encore en attente de leur proposition d'admission ont reçu cette proposition en moins de trois jours. Ils étaient 7 300 et 7 900 en 2021 et 2022.

Mais le CESP considère que cet effet serait encore plus significatif si la hiérarchisation s'effectuait aux alentours du 15 juin. Les candidats auraient le temps de voir le début des réponses à leurs vœux et pourraient être accompagnés dans leurs lycées. Les épreuves écrites terminales du baccalauréat 2024 se déroulant du 18 au 21 juin, on peut craindre que la date de la hiérarchisation soit plus tardive pour éviter un stress supplémentaire au moment des dernières révisions. Cependant, ce souci légitime ne prend pas en compte le stress de milliers de candidats qui attendront, de ce fait, encore plus longtemps leur proposition d'admission. »

Résumons…

Parcoursup est fondé sur un algorithme central irréprochable. Les procédures de classement local des formations sur lesquelles il s'appuie semblent socialement fondées et sont techniquement pertinentes, car compatibles avec l'AGS, même si le CESP note qu'elles doivent gagner en transparence. Mais, soit les pouvoirs publics font preuve d'impéritie en confondant le fait qu'il convient de ne pas utiliser les préférences des candidats dans les classements locaux avec le fait de ne pas les leur demander pour l'algorithme central. Soit ils organisent délibérément un simulacre d'interaction usager-algorithme au détriment des candidats chaque été.

Et le baccalauréat, dans tout cela ?

Pour donner à Parcoursup le temps de s'exécuter, le législateur a en 2018 une idée de génie : avancer au mois de mars les épreuves de spécialité du baccalauréat. On sait ce qu'il advint : au lieu de ne déserter le lycée qu'au mois de juin comme depuis le début des années 2000, les élèves quittaient le navire terminale dès la fin mars. Un tiers de la formation s'en allait ainsi à vau l'eau. Le MENJ s'en est ému dès le mois de juin 2023. Une proposition pas exactement à la hauteur des enjeux émerge d'abord[12] : cacher aux élèves et aux familles les notes des spécialités pendant trois mois. Peu après, le ministre de l'Éducation nationale Pap Ndiaye passe cette proposition par profits et pertes et brosse des

12. Pierre Mathiot, France Info, 14 juin 2023.

aménagements plus conséquents[13]. Lui emboîtant le pas, son successeur, Gabriel Attal, prend une décision frappée au coin du bon sens en replaçant les épreuves du baccalauréat en juin…

Parcoursup : 1 – Baccalauréat : 0

Survient alors une nouvelle difficulté : comment, avec des épreuves aussi tardives, donner à Parcoursup le temps de s'exécuter ? En n'intégrant à Parcoursup que les résultats du français et de la spécialité abandonnée en fin de première pour une portion congrue, et les résultats aux contrôles continus, pour l'essentiel. En bref : un stress continu des élèves durant l'année à chaque contrôle continu. Un stress qui… continue durant les pérégrinations estivales de Parcoursup. Comment échapper à ce hiatus ? Par des épreuves de spécialité si anticipées qu'elles obèrent un tiers de la formation de la terminale, mais qui peuvent être intégrées à Parcoursup ? Ou alors par des épreuves à leur place naturelle — à la fin de la formation — mais qui rendent les résultats du baccalauréat inexploitables pour Parcoursup ? Les pouvoirs publics ont tranché : priorité à Parcoursup. Les résultats du baccalauréat brillent par leur inutilité, encore accrue par le fait qu'en moyenne, un seul élève par classe échoue au baccalauréat général.

Une autre voie ?

Il existe pourtant une autre voie : ré-interroger, comme le fait le CESP en 2023 et implicitement en 2024, le fondement du dogme imposant aux élèves de ne pas hiérarchiser leurs vœux. Avec APB, la hiérarchisation était totale. Chaque élève pouvait effectuer jusqu'à 24 vœux, tous hiérarchisés. Avec Parcoursup, toute hiérarchisation est théoriquement bannie. Position dont il était évident en 2018 qu'elle était un principe érigé en dogme.

Entre ces deux extrêmes, dessinons une voie médiane, déjà esquissée par Claire Mathieu en 2022[14], qui consisterait à ne pas interdire à l'élève de hiérarchiser ses vœux, ni à l'y obliger. Mais à le lui permettre. Totalement. Ou partiellement. Ou par blocs. Sur 10 vœux, il pourrait hiérarchiser, par exemple, les trois premiers. Ou les cinq derniers. Ou, autrement dit, déclarer certains vœux ex-æquo. Ou encore indiquer des formations vivement souhaitées, souhaitées, potentiellement souhaitées, par blocs. Il y a fort à parier que nombre de candidats auraient à cœur de hiérarchiser leurs vœux. Parce que c'est humain. Parce que c'est rassurant. Parce qu'il est naturel d'avoir des préférences entre plusieurs formations souhaitées, que ces préférences soient rationnelles, changeantes, ou non. Modifier encore l'expression des vœux engendrerait sans doute

13. Pap N'diaye, France 2, 22 juin 2023.

14. Parcoursup : « La non-hiérarchisation des vœux est un sujet conflictuel qui pourrait être enterré par l'adoption d'une solution médiane », Le Monde, tribune de Claire Mathieu, 11 octobre 2022.

un stress nouveau au sein des familles de candidats, de nouvelles explications à dispenser aux candidats eux-mêmes dans les lycées. C'est ce qui explique notre proposition de rendre la hiérarchisation facultative et partielle. En 2024, sans hiérarchisation initiale [graphique-1] 533 000 candidats sur 946 000 ont eu une proposition positive. 5 jours plus tard ils étaient 91 000 candidats de plus, et encore 86 000 après 10 jours. 710 000 candidats au total ont donc eu une réponse au bout de 10 jours. Les 236 000 restants attendront le 12 juillet, ou la phase complémentaire. Et avec une hiérarchisation ? Une simulation basée sur les vœux de 2024, avec un ordonnancement total pour tous les candidats a été effectuée[15][graphique-1].

Le résultat est éloquent : 730 000 auraient eu une proposition dès le premier jour, soit 197 000 de plus que sans hiérarchisation. 20,8 % des candidats auraient donc une réponse instantanée au lieu d'attendre 5 à 10 jours. Et le nombre total de réponses instantanées est même un peu supérieur à celui obtenu actuellement en 10 jours. En réalité le gain serait moindre, car certains candidats ne hiérarchiseraient pas leurs vœux, et d'autres ne les hiérarchiseraient que partiellement[16], mais les gains obtenus depuis deux ans par hiérarchisation tardive et la simulation précitée sont prometteurs.

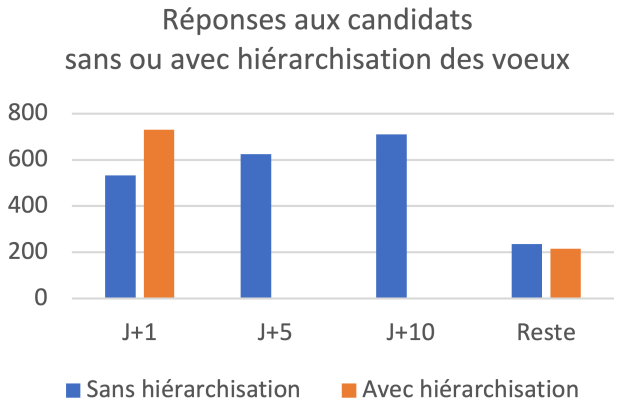


Fig. 1. Réponse aux candidats sans ou avec hiérarchisation des vœux.

15. La simulation d'ordonnancement effectuée par Hugo Gimbert tente de refléter les préférences réelles des candidats en se basant sur un modèle de popularité utilisé en économie. Le modèle est entraîné sur les choix des candidats effectués lors de la session 2023.

16. Il s'agirait de faire varier le taux de candidats qui hiérarchisent leurs vœux ainsi que pour chacun le taux de hiérarchisation et d'effectuer une simulation pour chaque cas, afin de déterminer comment varie le gain de temps selon la variation de ces deux taux, mais cette étude dépasse le cadre de cet article.

Faire tomber le dogme de la non-hiérarchisation

La difficulté principale ne réside cependant pas dans notre proposition. Techniquement, elle est triviale : elle est présente — en sommeil — dans le code de Parcoursup. Scientifiquement le MESR est armé pour la réveiller. Socialement, elle est raisonnable de par son caractère facultatif et partiel. Mais politiquement, elle nécessite de reconnaître officiellement que le dogme de la non-hiérarchisation des vœux a vécu… c'est le plus difficile ! En attendant cette reconnaissance du MESR, certains services opérationnels et les recommandations éclairées du CESP ont permis de faire naître la hiérarchisation en 2022, de l'avancer dans le calendrier 2023. Mais à ce jour l'avancement n'a pas franchi la date des épreuves du baccalauréat, malgré les propositions du CESP. Peut-être la vraie solution n'est-elle pas de grappiller quelques jours, année après année, mais simplement d'avancer la date de la hiérarchisation avant le lancement du processus d'appariement proprement dit. Cela impliquerait de faire tomber le dogme de la non-hiérarchisation afin que l'utilisation effective de Parcoursup ne s'étale pas au-delà de la période suivant immédiatement la proclamation des résultats du baccalauréat. Parcoursup pourrait rendre son verdict très rapidement pour l'écrasante majorité des candidats. Souvenons-nous que 95 % des candidats obtenaient une réponse avec APB en 2017 en une semaine, contre 79,3 % des candidats avec Parcoursup en 2023 pour la phase principale.

Et celui du nombre de vœux ?

Au passage… pourquoi limiter le nombre de vœux ? Pourquoi ne pas faire confiance aux candidats ? Ils en font plus sous Parcoursup en moyenne que sous APB, malgré la limitation plus forte. Officiellement, les candidats sont limités à 10 vœux, limite présentée comme une amélioration à priori par rapport à APB. Mais la mécanique de Parcoursup autorise de nombreux sous-vœux. En 2023, le nombre de vœux a été compris entre 1 et… 255. Un peu moins d'un million d'étudiants a effectué plus de 16 millions de vœux, soit 17,6 vœux en moyenne. À l'heure ou Parcoursup est sur le point de dépasser l'âge auquel APB fut mis au rebut, pourquoi ne pas le libérer du dogme du nombre de vœux, puisqu'en pratique, ils ne sont pas limités ? Le seul « risque » que cela ferait prendre à Parcoursup serait qu'il s'exécute d'autant plus vite que les listes des vœux seraient ordonnées.

Parcoursup instantané, non. Encore plus rapide, oui !

Deux limites à l'exécution instantanée de Parcoursup sont cependant indépassables : d'une part, les candidats formulent plusieurs vœux. Ils finissent par en accepter un seul, libérant des places qu'ils avaient préemptées, au bénéfice d'autres candidats. D'autre part un nombre significatif de candidats renoncent à

toute poursuite d'études supérieures et sortent de Parcoursup, libérant tardivement de nouvelles places. L'attente inhérente au processus peut expliquer une partie de ce nombre, plus important qu'avec APB. En outre de nombreux étudiants, qui craignent d'avoir échoué à leur année de L1, sont candidats sur Parcoursup, puis libèrent leur place après avoir obtenu un résultat positif à leur année. Ces deux raisons impliquent que Parcoursup se déroule, comme APB, en plusieurs phases. En six années d'améliorations continues, le temps de réponse de Parcoursup a été divisé par presque trois. Il reste à circonvenir le dogme de la hiérarchisation des vœux et au passage celui — certes secondaire — du nombre des vœux pour accélérer encore les réponses de Parcoursup et limiter le stress d'un nombre croissant de candidats pour rendre instantanée ou presque la première et principale phase de réponse.

Des indicateurs de satisfaction à la baisse depuis la réforme du lycée

Les nombreuses améliorations déjà apportées à l'application[17] n'ont pas suffi à juguler ce stress : le taux de candidats stressés par Parcoursup est passé de 77 % à 83 % entre 2020 et 2023[18]. Sa clarté, sa fiabilité et sa personnalisation demeurent bien perçues mais elles baissent toutes entre 2020 et 2023 [graphique-2]. Les avis positifs sur sa rapidité, sa transparence et son équité chutent entre 2020 et 2023, se rapprochant ou passant sous les 50 %, avec un inquiétant 31 % pour l'équité.

La présentation en trompe-l'œil du MESR, restreignant le propos à l'année antérieure ou aux indicateurs positifs mais sur des champs secondaires (comme la visibilité des médias utilisés par Parcoursup) n'y change rien. Entre 2020 et 2023, de nombreux indicateurs humains sont à la baisse : de 82 % à 73 % pour l'accompagnement à l'orientation par le professeur principal, de 82 % à 76 % pour l'accompagnement à l'utilisation de Parcoursup, de 56 % à 43 %[19] pour la formulation et la confirmation des vœux.

Un sommet est atteint à propos de la hiérarchisation des vœux. Il consiste à demander aux élèves si « le fait de ne pas classer au mois de mars les vœux de poursuite d'études que l'on formule » leur a plu : 73 % répondent positivement. Mais le fait de « pouvoir formuler des vœux de poursuite d'études librement sans avoir à les classer par ordre de préférence » recueille, lui, 31 % d'intérêt. Il est inférieur à l'intérêt d'être fixé sur ses classements avant le bac (34 %) et très inférieur à celui de connaître son classement et les chances d'intégrer l'une des formations choisies en cas de placement en liste d'attente (45 %).

17. Pour 2024, l'accessibilité aux personnes en situation de handicap, l'information sur l'insertion professionnelles, la comparaison d'indicateurs des formations.
18. Baromètre Parcoursup 2023, Enquête CSA, septembre 2023.
19. De 2021 à 2023 ; donnée non disponible pour 2020.

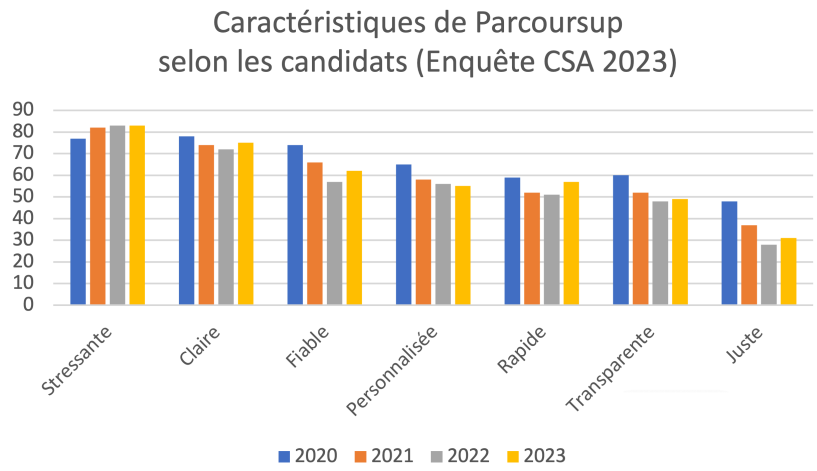


Fig. 2. Caractéristiques de parcoursup selon les candidats (Enquête CSA 2023).

Parcoursup en danger ?

Aujourd'hui, Parcoursup semble durablement installé dans le paysage de la formation secondaire et supérieure français. Mais nombre d'indicateurs sont à la baisse depuis que les enquêtes de satisfaction ont débuté avec la réforme du lycée en 2020. L'évaluation auto-satisfaite du MESR, en contradiction même avec les résultats des enquêtes qu'il commandite, le positionnement opportuniste de personnalités politiques à la peine pour capter l'attention d'un électorat jeune et quelques inconvénients résiduels réels de Parcoursup pourraient avoir raison de lui aussi vite qu'ils ont eu raison d'APB.

Une proposition pragmatique

Minimiser significativement le stress des candidats en réduisant leur attente par accélération du processus Parcoursup devrait permettre d'enrayer cette chute et de préserver Parcoursup d'un sort funeste à court terme. En outre, avec un calendrier adapté, les résultats du baccalauréat seraient connus avant le lancement de Parcoursup. Les épreuves terminales et le baccalauréat lui-même conserveraient alors quelque lueur d'intérêt, et au passage, le stress continu engendré par l'importance actuelle du contrôle continu diminuerait. Pragmatique, notre proposition consiste donc à introduire dans Parcoursup une dose officielle, paramétrable par le législateur ou par l'élève lui-même, de hiérarchisation des vœux pour espérer que près des deux-tiers des candidats obtiennent

une réponse dès le premier jour de la phase principale de Parcoursup. Nous ne saurions trop conseiller la prise en compte des rapports de l'OPECST et du CESP précités pour étayer cette proposition : ils contiennent en germe l'essentiel des analyses que nous effectuons.

La hiérarchisation des vœux à la carte ne nécessite que des adaptations informatiques marginales. Son intérêt est aisément explicable aux élèves et à leurs familles. Son coût financier est négligeable. Son gain humain est immense. Puissent les ministères concernés prêter quelque attention à cette proposition de hiérarchisation sur mesure afin que l'accession aux études supérieures ne relève plus, pour les élèves et leurs familles, d'un parcours du combattant.

Nouveau lycée : « plus spécialisé que moi, tu meurs ! »

La SIF

Processus de choix complexe, erreurs d'orientation dans l'enseignement supérieur, surspécialisation précoce : tels sont les ingrédients de la cuisine « lycée général » à la sauce « bouquet de spécialités ». Il y a urgence : modifions la recette significativement et rapidement avant l'indigestion généralisée en simplifiant l'orientation, en informant sur les prérequis du supérieur, en élargissant le spectre de formation des élèves. Action !

Avant la réforme

De 1965 à 2019, les études au lycée général sont organisées par l'État sous forme de séries. Chaque série est définie comme un ensemble cohérent de disciplines s'articulant autour de plusieurs disciplines principales. Cette période se caractérise par la montée en puissance des disciplines scientifiques : en 2018, le baccalauréat S concerne 52 % des bacheliers.

La réforme et les spécialités

Pour faire pièce à la suprématie de la filière S et pour répondre à une hypothétique demande sociétale de profils diversifiés, la notion même de série vole en éclats en 2018, sous la férule de Jean-Michel Blanquer, ministre de l'Éducation nationale. Ce n'est plus l'État qui organise les études au lycée… C'est l'élève lui-même. Après quatre mois de classe de seconde, parmi treize possibles, il choisit trois spécialités qu'il étudiera en classe de première. En première, *bis repetita* :

DOI:10.48556/SIF.1024.24.49

il peaufine l'organisation de son cursus en choisissant celle de ses trois spécialités qu'il abandonnera en terminale. Fini le groupe-classe… Exit le professeur principal… Oubliée la cohérence interne à chaque série… Et vive les nouvelles inégalités sociales devant des choix de spécialités supposés ne pas conditionner celui des filières de l'enseignement supérieur ! Bravo aux choix incohérents ! Bienvenue aux regrets à l'orée des études supérieures !

Le tronc commun

En contrepoint de ce libre choix de spécialités surgit le tronc commun. Il représente plus de 55 % du volume horaire total, soit seize heures d'enseignement dont deux seulement sont dévolues aux sciences. Tous les élèves étudient obligatoirement deux langues vivantes, mais moins de 5 % font de l'informatique en terminale. Ils suivent quatre heures de français et autant de philosophie. Mais, avec deux spécialités seulement, plus aucun ne bénéficie d'un enseignement en terminale fondé comme avant la réforme sur l'approfondissement de trois disciplines. Ce tronc commun quasiment sans sciences et cet approfondissement insuffisamment polyvalent conduisent à ne proposer aux élèves qu'une vision du monde trop parcellaire. Il est inquiétant que les futurs citoyens ne bénéficient pas d'un enseignement leur garantissant d'acquérir les notions permettant de comprendre les enjeux du numérique, du réchauffement climatique ou des biotechnologies.

Les spécialités : concentration des choix…

La réforme a certes permis d'ériger en spécialités complètes des disciplines qui n'étaient auparavant qu'à faible volume horaire. Ainsi en va-t-il de l'informatique (*Informatique et sciences du numérique*, NSI) passée de deux heures en terminale S à quatre heures en première et six en terminale (*Numérique et sciences informatiques*, NSI). Mais cette spécialité n'est proposée que dans un lycée sur deux ; le nombre de postes au CAPES et à l'agrégation d'informatique demeure très faible. Parallèlement, les effectifs en « sciences de l'ingénieur » ont été divisés par près de quatre en cinq ans. Le but de la réforme était de diversifier les parcours des élèves. Mais en pratique, 44 % des élèves choisissent les trois mêmes combinaisons. Et sur les 78 combinaisons possibles, 68 sont choisies par moins de 3 % des élèves : la diversification effective des profils se fait attendre… d'autant plus que le resserrement des choix sur les spécialités principales s'accroît depuis deux ans. Cela s'explique par l'accroissement de choix paradoxaux : de nombreux élèves souhaitant faire des études d'informatique choisissent d'abandonner la spécialité NSI en terminale pour conserver *Physique-chimie*, au prétexte que cela leur ouvrirait plus de portes après le bac.

… et ajustements inopportuns

Ce sont les ajustements censés pallier l'abandon de spécialité en terminale. Eu égard à l'abandon d'une spécialité en terminale, même les spécialités les plus choisies le sont moins qu'avant la réforme. Les élèves « sans maths » ne représentaient que 13 % des effectifs avant la réforme. Ils sont 40 % depuis. En première, une rustine d'1 h 30 de mathématiques est ajoutée, obligatoire pour les élèves qui n'ont pas choisi la spécialité *Mathématiques*… en plus de leur emploi du temps standard. Les élèves qui abandonnent les mathématiques en terminale peuvent choisir l'option *Mathématiques complémentaires*. Ces ersatz sont insuffisants pour suivre un parcours en sciences ou en sciences économiques et sociales ; ils ferment les portes correspondantes de l'enseignement supérieur. À l'inverse, le programme de la spécialité *Mathématiques* est trop ambitieux pour les futurs économistes : près de 80 % d'entre eux abandonnent cette spécialité en terminale. Les *Sciences de l'ingénieur* ne peuvent être étudiées sans un minimum de physique… deux heures de physique obligatoires ont donc été ajoutées pour les élèves suivant cette spécialité. Si ces élèves suivent aussi la spécialité *Physique-chimie*, ils subissent en outre des enseignements partiellement redondants. Bref, plutôt que de rendre les parcours des élèves cohérents, les mesurettes maladroites ajoutées ne font que compliquer les parcours, sans apporter de réelles solutions.

L'addition se paie dans le supérieur

Choix de spécialités non coordonnés, élèves mal informés, professeurs sans formation appelés à guider les choix des élèves… le tout mâtiné d'un discours « tout choix de spécialité permet toute poursuite d'étude dans l'enseignement supérieur » : c'est le cocktail détonant qui conduit à de multiples erreurs d'orientation et à autant d'échecs dans l'enseignement supérieur. Par exemple, 30 % des élèves ayant suivi la spécialité *Numérique et sciences informatiques* ont abandonné la spécialité *Mathématiques*. Ils ne découvrent qu'en début de licence qu'un bagage en mathématiques non négligeable est indispensable pour réussir des études en informatique. Les licences sont contraintes de proposer des modules de remédiation ou des parcours en quatre ans, pour contrecarrer les conséquences de choix de spécialités inopportuns. Le coût engendré est doublement préjudiciable : humainement, pour les étudiants catalogués « oui, si » et financièrement pour les universités, déjà exsangues à cause des crises économiques récentes.

Trois spécialités en terminale

Une caractéristique centrale du nouveau lycée augmentant les échecs en première année postbac réside dans la surspécialisation précoce, due à

l'abandon de la troisième spécialité en terminale. Trois spécialités, c'est que ce proposaient peu ou prou les anciennes séries. L'abandon d'une spécialité en terminale, outre le manque de 60 % de la discipline concernée, conduit à effacer les notions correspondantes étudiées en première et seconde.

Aussi, les améliorations suivantes sont-elles indispensables pour éviter d'envoyer des cohortes d'étudiants, chaque année grandissantes, vers un échec assuré :

- stopper l'hypocrisie du tout à la carte et **définir un ensemble restreint de triplettes de spécialités cohérentes avec des choix d'orientation** ; par pragmatisme et faute de moyens, certains lycées pratiquent déjà cette restriction d'une part en n'offrant que quelques spécialités et d'autre part en contraignant les associations de spécialité ;
- **autoriser les établissements recrutant en postbac, y compris pour leurs formations non sélectives, à définir pour chaque formation des spécialités prérequises** pour offrir aux jeunes des chances significatives d'y réussir ;
- **conserver en terminale les trois spécialités choisies en première** (et revoir en conséquence, à la baisse, le volume du tronc commun).

Ces améliorations ne résoudront pas tous les problèmes posés par la réforme du lycée. Mais, en simplifiant le système, et associées à une simplification de Parcoursup, elles limiteront les inégalités sociales et de genre, et surtout le gâchis humain engendré par la multiplication des erreurs d'orientation et les déceptions afférentes dans l'enseignement supérieur. Il est plus que temps de rouvrir le débat sur la structure même du lycée !

La spécialité *Numérique et sciences informatiques* au lycée en 2024 : péril en la demeure !

La SIF

Dans le paysage entièrement recomposé du lycée général depuis sa réforme de 2019, la spécialité « Numérique et sciences informatiques » (NSI), est apparue à égalité aux côtés de spécialités plus classiques. Son avènement soufflait un vent de modernité sur le lycée, offrant de nouvelles possibilités de profils de formation pour les élèves. Il semblait en phase avec l'explosion du numérique et de l'informatique dans tous les pans de notre société, et précédait de peu la diffusion soudaine de l'« intelligence artificielle » auprès du grand public.
Quelle est, cinq ans plus tard, la situation de la spécialité NSI ? Décryptage, chiffres de la note n°24-06 de la DEPP[1] à l'appui.

Comme pour toutes les spécialités, des données sont disponibles sur les effectifs des élèves pour la spécialité NSI sur quatre ans. Dans une première analyse de cette spécialité [Bulletin 1024, n°21] nous mettions en évidence qu'elle était née à un niveau très bas, concernant à peine plus d'un élève par classe de terminale. Dans notre deuxième analyse [Bulletin 1024, n°22], nous relevions la poursuite d'une hausse timide, en notant qu'au rythme de progression observé, NSI atteindrait le taux moyen des spécialités en… 2045. Quid de la cuvée 2023 ? Une stagnation, voire un tassement, couplés à une absence totale de politique publique accompagnant sa mise en place, caractérisent l'évolution récente de la spécialité NSI.

1. Direction de l'évaluation, de la prospective de la performance.

DOI:10.48556/SIF.1024.24.53

NSI en terminale : des effectifs globaux en légère baisse et une proportion moindre dans le contexte d'un secteur sciences en baisse.

4,6 % des élèves de terminale ont choisi NSI en 2023. Soit 0,1 point de moins qu'en 2022 (−223 élèves), après une progression de 3,7 % à 4,7 % de 2020 à 2022. Ce tassement survient dans le cadre d'une augmentation globale de 1421 élèves et d'une progression du secteur sciences de 3,4 points depuis 2022, masquant des différences majeures : les mathématiques progressent de 4,2 points, et les autres sciences régressent au total de 0,8 point. Depuis 2020, le cas des SVT est le plus inquiétant avec un recul total de 4 points. Au total, sur quatre ans, le secteur sciences recule de 3,7 points tandis que les maths progressent de 2,6 points (figure 1).

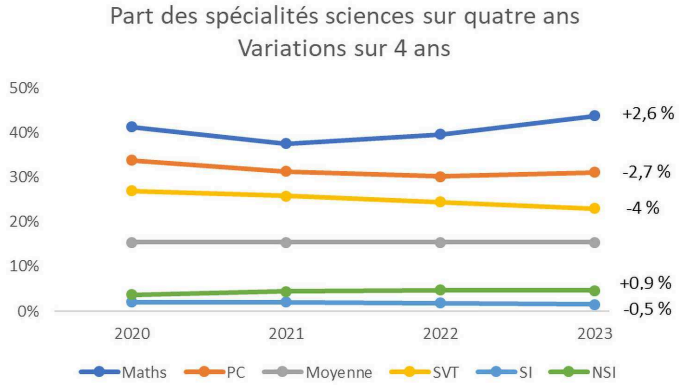


Fig. 1. Part des spécialités sciences et variations sur quatre ans.
(Part moyenne d'une spécialité en gris, à 15,4 % = 200 % / 13 spécialités.)

En 2024, les moins de 5 % des élèves de terminale qui suivent la spécialité NSI représentent à peine 1,4 élève par classe en moyenne. Et 44 % des lycées ne proposent pas la spécialité NSI en 2024… notamment parce que les effectifs attendus sont trop faibles, en particulier en terminale.

Mais limiter l'analyse aux seules années postréforme et à la spécialité NSI serait réducteur ; examinons cela.

– NSI progresse certes de 0,9 point entre 2020 et 2023 pour arriver à 4,6 % des effectifs, soit 17612 élèves. Ce chiffre est à mettre en regard des 23013 élèves qui suivaient l'option « Informatique et sciences du numérique » (ISN) en 2017, et des 18662 élèves qui la suivaient encore en 2019.

– ***En 2024, la spécialité NSI est, du point de vue des effectifs, à un niveau voisin de l'option ISN en 2014***. Et le fait que NSI marque le pas en proportion et en nombre d'élèves n'incite pas à l'optimisme : l'objectif, modeste, de retrouver l'effectif de 2019 paraît difficile à atteindre.
– ***La progression des mathématiques, toute relative, survient après une chute sans précédent*** au moment de la mise en place de la réforme. Un exemple parmi tant d'autres : avant la réforme, 100 % des élèves de la série ES suivaient 4 h ou 5,5 h de maths par semaine. Depuis la réforme, plus des trois quarts des élèves (75,9 %) suivant la spécialité SES ne suivent pas la spécialité maths en terminale, et seuls 18,3 % de ces derniers suivent l'option MC : cela signifie que plus de 62 % des élèves ayant choisi la spécialité SES ne font plus de maths du tout en terminale.
– 70 % des étudiants ayant choisi NSI en terminale suivent la spécialité maths, contre 63,5 % en 2021 et 67,4 % en 2022. Cette augmentation est directement liée à celle des effectifs de la spécialité maths. Parmi les 30 % d'élèves ayant choisi NSI sans la spécialité maths, seuls 52 % de ceux qui ont choisi l'une des spécialités PC, SVT ou SI choisissent l'option MC. Et parmi ceux qui ont choisi l'une des deux spécialités LLCER et SES (choisies plus souvent que les trois précédentes), seuls 30,2 % et 33,6 % d'entre eux suivent l'option MC. En bref, ***la majorité des élèves suivant NSI sans la spécialité maths ne suivent pas l'option MC non plus*** : ils ne font donc plus du tout de mathématiques en terminale, se fermant de fait la porte des formations du supérieur à dominante informatique, alors que par la structure même du lycée avant la réforme, 100 % des élèves qui suivaient l'option ISN faisaient des mathématiques.

NSI, une spécialité très abandonnée entre la première et la terminale

Si les effectifs et le taux d'élèves en terminale NSI reculent, ce n'est pas le cas en première, où NSI continue de croître légèrement. La raison de ces variations opposées ? Un taux d'abandon de NSI entre la première et la terminale, déjà massif avec 50,7 % en 2022, qui atteint 52,8 % en 2023, au profit des spécialités maths et PC.

En effet, les mathématiques sont passées en deux ans d'un taux d'abandon inquiétant (41,7 %) à un taux inférieur (32,7 %) au taux moyen (33,3 %). La spécialité PC bénéficie d'un taux d'abandon faible et stable (27,7 % entre 2022 et 2023), alors que la spécialité SVT suit un chemin inverse de celui des maths, passant en deux ans de 34,1 % à 39,1 % d'abandon. Avec la spécialité SI, perdant à présent 68,7 % des élèves entre la première et la terminale, la spécialité NSI fait figure de spécialité sinistrée. Elle dépasse très largement le taux moyen d'abandon, et même celui de la spécialité « latin / grec ». Les effets du recentrage

des élèves sur les deux spécialités, maths et PC, qui garantissent des taux de réussite élevés dans l'enseignement supérieur est patent.

Enfin, l'écart d'abandon entre filles et garçons demeure important : 62,5 % des filles abandonnent NSI en 2023 (60,9 % en 2022), contre 50,5 % des garçons, taux lui-même en hausse (48,4 % en 2022) (figure 2).

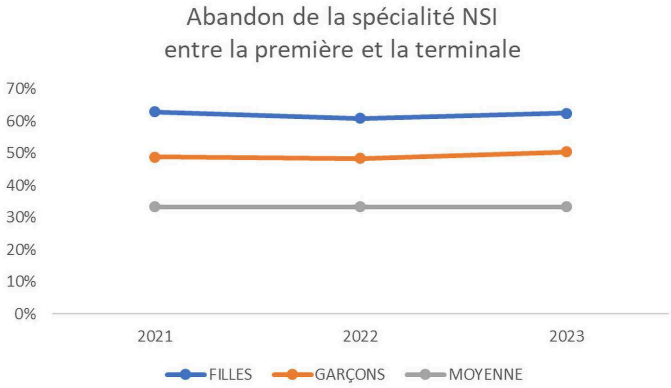


Fig. 2. Abandon de la spécialité NSI entre la première et la terminale. (Taux d'abandon standard à 33% : chute de 3 à 2 spécialités.)

Combien de filles en spécialité NSI ?

Le taux de filles en NSI a encore cru de 0,6 point, pour s'établir à 15,2 % en 2023 (figure 3). À ce rythme de progression, le plancher de 30 % fixé par le ministère en 2022 serait atteint non plus au début des années 2040 comme l'indique notre précédente analyse, mais plutôt vers 2050. Cela confirme que l'absence de politique volontariste, malgré la multiplication des dispositifs de communication à l'égard des filles en faveur du numérique, ne peut conduire à un taux acceptable de filles à un horizon proche. ***Le nombre de filles a quasiment stagné entre 2022 et 2023, passant de 2612 à 2673 filles, soit toujours de l'ordre de 1,5 fille par lycée***. Le taux de filles en NSI est d'autant plus faible que les filles représentent au total 55,6 % des effectifs en 2023.

La doublette maths-NSI ?

La doublette maths-NSI progresse à peine : elle représente 3,22 % des élèves en 2023, contre 3,15 % en 2022, soit 303 élèves de plus, à comparer à la progression de 9919 élèves pour la doublette maths-PC et à la progression de 16499 élèves en spécialité maths. Seuls 1,8 % des élèves supplémentaires en spécialité maths

choisissent NSI comme autre spécialité alors que 60,1 % de ces élèves choisissent la spécialité PC : la reconstitution de la série S — en version appauvrie par la perte d'une spécialité — semble « en marche », alors que le but premier, affiché, de la réforme du lycée était de faire pièce à sa suprématie en termes d'effectifs (52 % des élèves étaient concernés).

La proportion de filles poursuit sa pénible progression, passant de 11,4 % à 12 %, ce qui est un niveau plus bas encore (de plus de 3 points) que le taux de filles suivant NSI et une autre spécialité. 1478 filles suivent cette doublette en 2023, soit environ une fille par lycée offrant la spécialité NSI. Ces effectifs sont et demeurent confidentiels. Ils sont par exemple à comparer aux 29060 filles qui suivent la doublette maths-PC, soit 37,2 % des effectifs en 2023 contre 35,9 % en 2022, ce qui est une progression notable puisque le nombre d'élèves au total suivant cette doublette est passé de 17,9 % à 20,4 % sur cette période.

La triplette maths-PC-NSI ?

La triplette maths-PC-NSI poursuit sa lente progression, pour arriver à 5,3 % des effectifs en 2023 (20342 élèves). La proportion de filles progresse également un peu, s'établissant à 14,7 % (2994 filles). Cette proportion est à comparer d'une part à celle de la doublette maths-PC : 37,2 % de filles, et celle de la triplette maths-PC-SVT : 57,3 % de filles. En outre, toutes les triplettes contenant NSI et une autre spécialité scientifique sont très minoritaires en filles : la spécialité NSI agit donc bien comme un repoussoir pour les filles.

Après quatre ans d'existence, la spécialité NSI plafonne, au point de se tasser en terminale, à un niveau très bas. Actuellement, au lycée général :

Existe-t-il des indicateurs positifs de l'évolution future de la spécialité NSI ? Oui, mais ils sont trop ténus pour conduire à des taux satisfaisants à moyen terme. Par exemple, la part des élèves passe de 9,8 % à 10,1 % entre 2022 et 2023 en première… mais le taux moyen d'une spécialité est de 23,1 %. La proportion de filles en terminale passe de 13,1 % à 15,2 % entre 2020 et 2023… mais la proportion de filles au lycée est de 55,6 %.

Les effectifs de la spécialité NSI correspondent-ils à la place centrale qu'a pris le numérique dans la société ? Non. La proportion de filles répond-elle aux enjeux sociétaux d'égalité des chances d'accès aux emplois bien rémunérés et valorisés du numérique ? Non. En tant que spécialité scientifique, NSI soutient-elle la comparaison avec les spécialités scientifiques « historiques » ? Pas plus.

Il existe pourtant d'innombrables opérations de sensibilisation à l'informatique, en particulier à destination des femmes. Locales ou nationales, elles demeurent inopérantes, au regard des chiffres caractérisant NSI. L'aide à l'orientation ne semble pas plus efficace. En revanche, le changement de structure du lycée a induit, lui, des changements drastiques et soudains : baisse du nombre d'élèves scientifiques, et en particulier de celui des filles, baisse du

nombre et du taux de filles en informatique, augmentation des différences entre catégories socio-professionnelles, inégalités persistantes de territoires. La faible demande en postes NSI de la part des établissements a bon dos. L'absence de NSI dans un tiers des lycées et le faible nombre de postes au CAPES NSI et à l'agrégation lui sont trop aisément imputés.

Environ 90% des élèves ne suivent pas NSI en 1re	Plus de 95% des élèves ne suivent pas NSI en terminale	50,5% des garçons et 62,5% des filles abandonnent NSI après la 1re
1,4 garçon par classe suit la spécialité NSI	1,5 fille par lycée suit la spécialité NSI	Les filles représentent 15% des effectifs de NSI
Les effectifs NSI augmentent de 1,8% en 1re et baissent en terminale	La doublette maths-NSI stagne	Presque aucun nouvel élève choisissant la spécialité maths ne choisit NSI

Seule une politique publique agissant sur la structure du lycée, appuyée par un volet « ressources humaines » significatif pourrait pallier à court terme les travers de la réforme, qu'ils soient généraux, spécifiques aux sciences, ou plus spécifiques encore à la spécialité NSI. Pour la structure, sont en cause la quasi-absence de sciences dans le tronc commun et l'obligation d'abandonner une spécialité en terminale. Pour les ressources humaines, les quelques dizaines de postes au CAPES NSI et à l'agrégation d'informatique ne répondent pas aux besoins des établissements. Sans politique publique répondant à ces deux questions, NSI restera très abandonnée en terminale — surtout par les filles — et avec des effectifs demeurant sans rapport avec les enjeux socio-économiques.

Imaginons simplement que le nombre de garçons (14939) en première NSI représente, comme pour l'ensemble du lycée général, 44,4% des effectifs. Imaginons en outre que les spécialités de première soient conservées en terminale. On aboutirait ainsi à 70820 élèves en terminale NSI, soit un peu plus de quatre fois plus que les effectifs réels (17612 élèves) de 2023. Les effectifs de NSI seraient de fait d'un ordre de grandeur comparable aux effectifs des principales spécialités scientifiques. Mais, surtout, ils irrigueraient bien plus et mieux qu'actuellement cursus informatiques des universités, des classes préparatoires et écoles d'ingénieurs, et permettraient à court terme de répondre aux besoins socio-économiques et à la justice sociale dans le numérique. Seule une volonté politique forte permettrait de s'approcher de ces chiffres… existera-t-elle un jour ?

Un nouveau moteur pour 1024

Nicolas Taffin[1], Hervé Le Crosnier[1]

À partir de ce numéro – que vous avez entre les mains – le bulletin de la SIF, 1024, sera conçu et réalisé par C&F éditions. Cet article relate la refonte par cet éditeur de la chaîne éditoriale de 1024, depuis la réception des articles fournis par leurs auteurs jusqu'à la production d'un document prêt à être imprimé. L'approche utilisée est innovante et ouverte, basée sur les technologies du Web (HTML2Print), et son intégration dans la vie du comité éditorial de 1024 est faite sans solution de continuité afin de faciliter le travail fourni par l'équipe de bénévoles.

C&F éditions est une maison d'édition indépendante qui existe depuis 21 ans et publie principalement dans les champs de la culture numérique, de l'éducation et des communs de la connaissance. Depuis le début, nous nous y posons la question des outils de production du livre, et plus généralement des divers types de document. Curieux par nature, nous avons dès le début créé un service de R&D pour expérimenter diverses technologies, notamment celles liées aux outils du Web (HTML, CSS, Javascript, XML… de préférence avec des outils logiciels libres). Depuis 2018, nous pratiquons la mise en page de livres complets en HTML structuré et CSS[2]. Les progrès que nous faisons (pas à pas) nous ont permis de nous émanciper des logiciels propriétaires qui ont un quasi monopole dans l'industrie éditoriale.

1. C&F éditions.
2. Lire le chapitre « L'édition libre de C&F » qu'Antoine Fauchié consacre à C&F éditions dans sa thèse, *Fabriquer des éditions, éditer des fabriques. Reconfiguration des processus techniques éditoriaux et nouveaux modèles épistémologiques*, 2024, université de Montréal. https://these.quaternum.net/chapitre-05/#53-l%C3%A9dition-libre-de-cf.

DOI:10.48556/SIF.1024.24.59

Un désir de rencontre

En 2022, la rencontre avec la SIF s'est faite autour de l'idée d'un partenariat destiné à offrir aux lecteurs de *1024* des livres comme *Les entretiens de binaire* ou *Qui a hacké Garoutzia ?*, dont les auteurs sont également membres fondateurs de la SIF (ces ouvrages étant d'ailleurs réalisés en HTML2Print). Et quand l'équipe de *1024* autour de Denis Pallez nous a demandé conseil sur une réorganisation du flux de production de la revue, nous étions motivés. Cela permettait en outre de concrétiser le désir d'un membre de notre équipe, Yann Trividic, artiste, informaticien et lecteur assidu du bulletin, qui était à ce moment là en alternance chez nous pour son master d'édition. Restait à se mettre au travail.

Le problème, comme pour beaucoup de revues, est de s'affranchir de la grande hétérogénéité de formats (techniques et rédactionnels) des contributions. Cela demande beaucoup de temps d'harmoniser, d'amender, de préparer techniquement et de mettre en forme tous les textes et figures, et laisse moins de temps au travail des idées, la revue paraissant semestriellement. L'ouverture, la possibilité d'évoluer, figuraient en bonne place parmi les prémices de mise en place de ce nouveau flux éditorial, ainsi que la volonté d'intégrer plusieurs formats d'entrée.

Une architecture élaborée en discutant

Nicolas Taffin venait alors de terminer avec Julien Taquet et Agathe Baëz la chaîne éditoriale permettant les éditions en *open access* multiformat de catalogues scientifiques du musée du Louvre. Le premier ouvrage consacré au peintre Van Dyck était ainsi tout juste disponible dans plusieurs éditions réalisées à partir de la même source HTML (imprimé, site Web, pdf en ligne et epub[3]). Les besoins et le format de ce dernier projet étaient certes différents de la confection d'une revue (avec un sommaire et un rubriquage vivant au fil des contributions, les allers-retours entre éditeurs et auteurs, son fameux bouclage et ses contraintes, comme le DOI — identifiant d'objet numérique[4]). Mais des jalons étaient posés, et il y eut ainsi moins d'appréhension à se lancer.

Nicolas proposa un scénario reposant sur une architecture technique adaptée aux habitudes du bulletin, allant de la conversion de fichiers Word et Latex à la publication Web et print en passant par la sélection et les corrections. Il opta pour un format lisible et convertible, facilitant ces aspects (Asciidoc). Et Yann se lança courageusement dans les premières expérimentations : trouver les briques logicielles pour compléter les savoir-faire déjà maîtrisés, et tester différentes possibilités de mise en pages adaptées à une revue (choix des caractères, lisibilité, identité).

3. https://livres.louvre.fr.
4. https://www.doi.org.

Une chaîne de collaboration

Examinons maintenant plus en détail les principes qui nous ont guidés pour cette réalisation, puisque la SIF montre la voie à de nouveaux usages et usagers pour desserrer plus encore l'étau des logiciels propriétaires sur les productions graphiques.

Séparer le contenu de la mise en forme

La logique de fond de notre travail consiste à séparer la structure et le contenu du document de ses différentes incarnations, qui ne sont que différentes mises en forme d'une même source. Pourquoi ? D'abord parce que cela permet de travailler mieux, de se partager le travail, et aussi parce que cela rend une publication plus intelligente (la sémantique pilote explicitement la mise en forme), pérenne (formats ouverts) et interopérable (compatible avec de nombreux outils). Mais un bénéfice important de cette approche est aussi de pouvoir imaginer différents supports pour une même publication, et de les mettre en œuvre relativement simplement, ces différentes formes étant toutes des représentations d'un même document source. On appelle cela les publications multiformat ou à source unique *single source of truth*. Ce modèle renforce l'idée d'utiliser HTML / CSS, dont la séparation est au fondement des concepts, des normes et des outils. Dans cas du *single source* en HTML et CSS, cela signifie en gros que l'on utilise différentes feuilles de style pour un même document. « En gros », car en réalité c'est un peu plus complexe que cela[5].

Des traditions de composition

Changer ses habitudes est toujours un peu compliqué, et en premier lieu, pour nous-mêmes. La PAO (publication assistée par ordinateur) repose sur deux branches dominantes :

– Adobe InDesign, qui s'inscrit dans la tradition du WYSIWYG[6] popularisée par PageMaker et QuarkXPress. Un logiciel propriétaire complet et précis, mais au format fermé et au modèle «*cloud*», qui contraint au renouvellement régulier de l'abonnement sans lequel on ne peut plus accéder à ses fichiers. Les variantes WYSIWYG libres, comme Scribus, fonctionnent mais ne s'émancipent pas des limites inhérentes à ce modèle ;
– Latex qui excelle dans la composition, est libre, avec un modèle WYSIWYM[7] assez différent mais dont l'usage commun est l'emploi de classes ou modèles

5. Nicolas Taffin, « les illusions de la source unique », in *Typothérapie*, C&F éditions, 2023. En ligne : `https://polylogue.org/les-illusions-de-la-source-unique/`.
6. *What You See Is What You Get*, ou tel écran, tel écrit (une fois n'est pas coutume, la traduction en français plus compacte).
7. *What You See Is What You Mean*, ce que vous voyez est ce que vous voulez dire. `https://fr.wikipedia.org/wiki/What_you_see_is_what_you_mean`.

préétablis ou l'ajout d'une quantité de greffons qui rendent complexe l'interopérabilité et l'exploitation du code source. En outre, Latex, par sa nature même, incite à incorporer des instructions de mise en forme au contenu lui-même (au lieu de les séparer). Malgré sa précision microtypographique, Latex n'a pas été adopté par la communauté créative du design graphique.

Depuis les débuts de la composition numérique, d'autres normes ont été développées pour encoder les textes qui devaient devenir des livres. La majeure partie des langages utilisés (SGML, XML…) sont des langages à balises. Cela permet d'une part de traiter des quantités considérables de signes, de maîtriser directement le code, à la différence des traitements de texte, et de jouer pleinement de l'emboîtage des blocs pour les déposer sur la surface de rendu graphique (écran, imprimante…). Ce faisant, il s'agit de systèmes WYSIWYM contrairement aux traitements de texte et logiciels de mise en page qui ont gagné la faveur des auteurs ou des graphistes.

Avec une exception : le Web. Les personnes qui pratiquent le Web voient en lui une autre voie : séparation du contenu et de la mise en forme, universalité du format, puissance logicielle, précision du stylage, interactivité et une assez bonne lisibilité du code (qui peut être améliorée avec des variantes légères du balisage). Tout est presque là pour une PAO du futur. *Presque* : nous travaillons dans cette direction en tout cas. La route est longue, des imperfections existent, bien sûr, mais les progrès sont assez rapides et on peut se persuader que c'est une voie qui vaut la peine de s'y investir… ce que nous avons fait.

HTML

Faire le choix d'HTML ou de Latex, revient donc à utiliser une logique de formateur de texte : un code accessible avec un éditeur de texte qu'il faut interpréter pour obtenir un rendu visuel. HTML et les technologies associées (DOM — Document Object Model —, feuilles de style CSS, Javascript) sont à la fois simples (dans le nombre d'éléments à mémoriser et dans les logiques d'organisation du texte) et largement ouvertes à des programmes complexes de manipulation du document. Les navigateurs Web, s'appuyant sur un mécanisme de normalisation permanent et actif sous l'égide du W3C, sont devenus des outils rapides, puissants, adaptés à tout type de présentation sur écran. Ils peuvent donc être utilisés comme moteurs de la mise en page utilisant HTML.

Notamment, l'interface DOM permet à des scripts de modifier le rendu du code HTML directement au sein du navigateur sans avoir à modifier le document source en HTML.

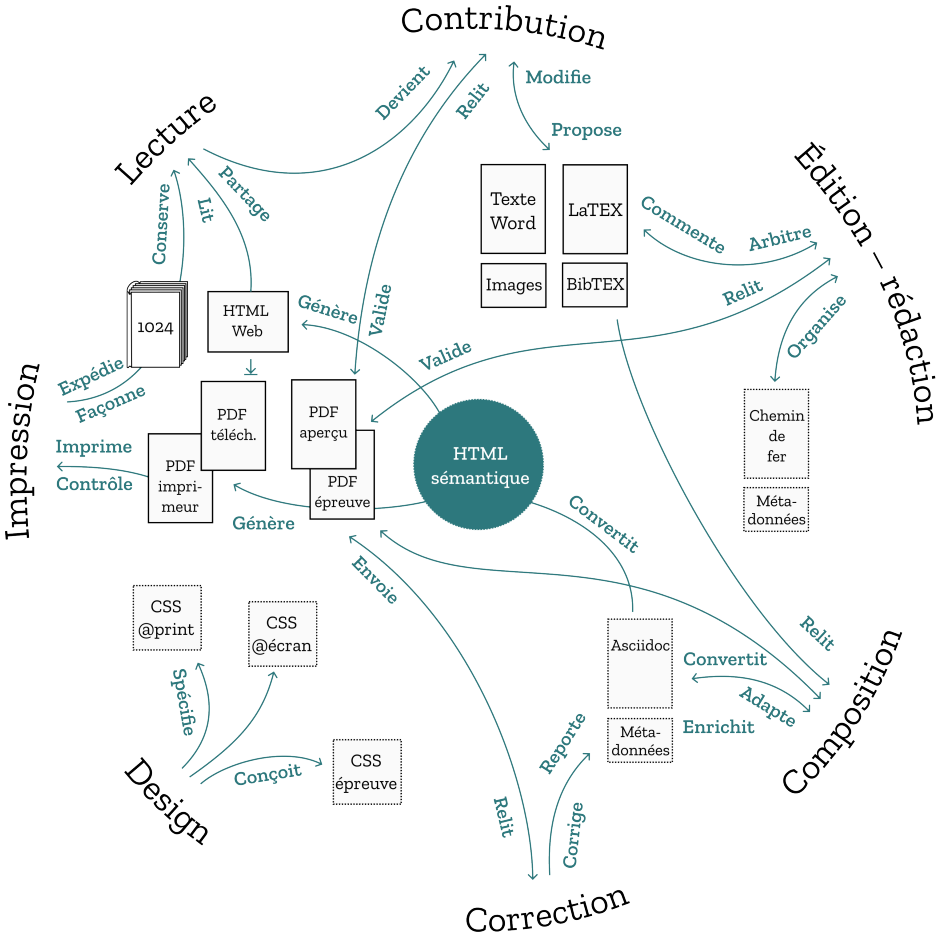


Fig. 1. Prendre en compte les échanges qui font la vie d'une publication. Ce graphique représente les rôles ou des casquettes, une même personne peut endosser plusieurs d'entre eux, mais dans l'idéal chacun apporte ses compétences.

Asciidoc, langage pivot pour rédiger le code HTML

HTML a des avantages et des inconvénients. Depuis sa création par Tim Berners-Lee pour échanger des documents, il est très répandu, très documenté, très outillé. Il a cependant trois défauts gênants pour l'édition : il est limité, il n'est pas fait pour l'imprimé, et il devient rapidement difficile à lire et écrire.

HTML est limité, il a relativement peu de structures. La définition de type de document HTML comprend des structures de base comme les titres, paragraphes, citations, quelques types de listes, du code, et des éléments interactifs (liens, formulaires). En édition, on peut avoir besoin de plus d'éléments, comme des exemples, des encadrés de différent types, des paragraphes spéciaux (chapeau, surtitre ou rubrique), des notes de bas de page, etc. Heureusement, cette limitation peut être outrepassée avec l'utilisation de classes, mais en la matière, il n'y a pas de standard, on passe donc en mode « personnalisé », chaque site Web développant ses propres classes (et les feuilles de style associées). Une fois les éléments fondamentaux définis, on prend en charge leur apparence visuelle à l'aide des feuilles de style CSS ou en manipulant le DOM. Le langage XML a essayé de dépasser cette limite, faisant en sorte que chaque fichier XML suive un modèle prédéfini, enregistré dans un schéma ou une DTD et accessible en ligne. Toutefois, les besoins sont souvent très différents et ce modèle devient rapidement assez lourd pour des usages ponctuels ou spécifiques, notamment des sites Web interactifs. Le vocabulaire de la TEI (*text encoding initiative*[8]) a été développé comme un outil généraliste pour décrire toutes les variantes de texte, avec beaucoup de précision. Il s'accompagne cependant d'une certaine complexité de mise en œuvre, ce qui le réserve encore aux éditions scientifiques en sciences humaines (l'excellente chaîne de production Metopes, utilisée par des presses universitaires s'appuie sur ce modèle[9]).

HTML n'est pas fait pour l'imprimé. Le Web déroule des documents comme des *rotulus*,[10] dans des formats d'écran très variés, dépendant des ressources de ses « clients », ce qui ne cadre pas tellement avec le degré de contrôle que demande une véritable édition imprimée sur une double page dans un format bien défini. Alors que le Web (et sa déclinaison dans le format ePub) adapte la mise en page à l'outil de lecture, l'histoire de la typographie imprimée cherche au contraire à ce que le lecteur puisse voir exactement ce que l'éditeur à organisé sur la page. C'est cette distinction qui rend difficile l'adoption d'une norme pour les médias paginés, malgré le travail du W3C en ce sens.[11] Pour adapter précisément le résultat à la fixité du média paginé, il faut donc en complément des styles définis par CSS,

8. `https://tei-c.org/`.

9. `https://mrsh.unicaen.fr/pluridisciplinaire/poles-pluridisciplinaires/pole-document-numerique/metopes/` et `https://www.metopes.fr/metopes-methodes.html`.

10. Rouleau de parchemin qui se lit dans le sens de l'enroulement à la différence du *volumen* qui se lit perpendiculairement à son enroulement.

11. `https://www.w3.org/TR/CSS2/page.html`.

modifier *a posteriori* le DOM produit pour répondre aux besoins de l'imprimé. C'est avec l'aide de Javascript que le média paginé pourra être préparé et mis en forme, principalement avec le script libre et ouvert PagedJS, et plusieurs modules qui sont ajoutés dans la chaîne de traitement.

HTML devient vite un peu compliqué à lire et écrire. Un document HTML est composé de texte et de balises symétriques (ouvrantes et fermantes), ainsi que de différentes entités. Le besoin de découper précisément le texte au moyen de ces balises éloigne le code source HTML du texte tel qu'il sera rendu visuellement. Cela peut finir par être gênant, quand on écrit ou quand on relit son texte. Peu de gens *écrivent* directement en HTML. Des initiatives d'allègement du balisage existent depuis longtemps, comme Markdown, créé par John Gruber, avec l'aide d'Aaron Swartz en 2004. Markdown permet d'écrire du code HTML qui ressemble beaucoup à du texte simple, et sera pourtant interprété et transformé en bonne et due forme, lors d'une conversion en HTML. Mais Markdown reste sommaire et ne couvre qu'une partie de ce que l'on peut faire en HTML. C'est normal et c'est même sa qualité : il se concentre sur l'essentiel nécessaire à un blogueur. Un autre format plus expressif est Asciidoc, qui est inspiré de Docbook (un vocabulaire XML) dont il reprend l'essentiel des caractéristiques. Asciidoc permet d'écrire de la documentation technique et incorpore beaucoup plus d'éléments que Markdown, comme des exemples, des tableaux, des notes, une bibliographie, etc. Il est en outre en train de devenir une norme.

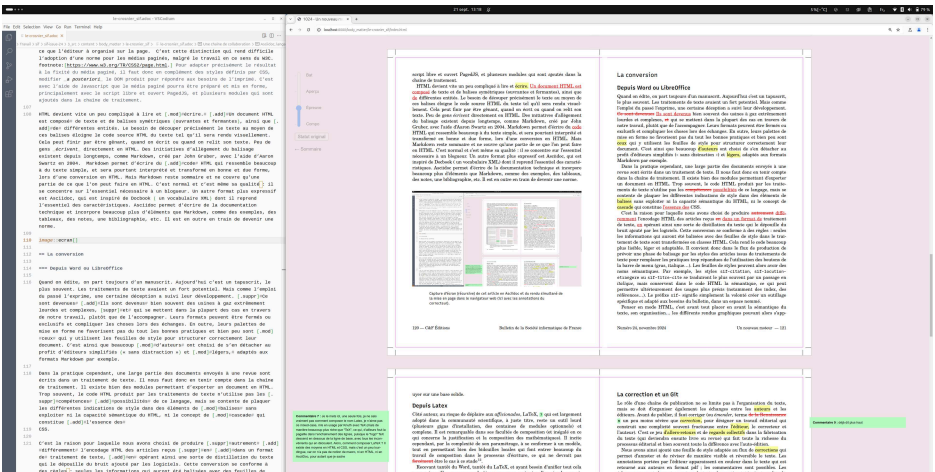


Fig. 2. Capture d'écran (récursive) de cet article en Asciidoc et du rendu simultané de la mise en page dans le navigateur web (ici avec les annotations du correcteur).

Le traitement des contributions

Depuis Word ou LibreOffice

Quand on édite, on part toujours d'un manuscrit. Aujourd'hui c'est un tapuscrit, le plus souvent. Les traitements de texte avaient un fort potentiel. Mais comme l'emploi du passé l'exprime, une certaine déception a suivi leur développement. Ils sont devenus bien souvent des usines à gaz extrêmement lourdes et complexes, qui se mettent dans la plupart des cas en travers de notre travail, plutôt que de l'accompagner. Leurs formats peuvent être fermés ou exclusifs et compliquer les choses lors des échanges. En outre, leurs palettes de mise en forme ne favorisent pas du tout les bonnes pratiques et bien peu sont ceux qui y utilisent les feuilles de style pour structurer correctement leur document. C'est ainsi que beaucoup d'auteurs ont choisi de s'en détacher au profit d'éditeurs simplifiés (« sans distraction ») et légers, adaptés aux formats Markdown par exemple.

Dans la pratique cependant, une large partie des documents envoyés à une revue sont écrits dans un traitement de texte. Il nous faut donc en tenir compte dans la chaîne de traitement. Il existe bien des modules permettant d'exporter un document en HTML. Trop souvent, le code HTML produit par les traitements de texte n'utilise pas les possibilités de ce langage, mais se contente de plaquer les différentes indications de style dans des éléments de balises sans exploiter ni la capacité sémantique du HTML, ni le concept de cascade qui constitue l'essence des CSS.

C'est la raison pour laquelle nous avons choisi de produire différemment l'encodage HTML des articles reçus dans un format de traitement de texte, en opérant ainsi une sorte de distillation du texte qui le dépouille du bruit ajouté par les logiciels. Cette conversion se conforme à des règles : seules les informations qui auront été balisées avec des feuilles de style dans le traitement de texte sont transformées en classes HTML. Cela rend le code beaucoup plus lisible, léger et adaptable. Il convient donc dans le flux de production de prévoir une phase de balisage par les styles des articles issus de traitements de texte pour remplacer les pratiques trop répandues de l'utilisation des boutons de la barre de menu (gras, italique…). Les feuilles de styles peuvent alors avoir des noms sémantiques. Par exemple, les styles `sif-citation`, `sif-locution-etrangere` ou `sif-titre-cite` se traduiront le plus souvent par un passage en *italique*, mais conservent dans le code HTML la sémantique, ce qui peut permettre ultérieurement des usages plus précis (notamment des index, des références…). Le préfixe `sif-` signifie simplement la volonté créer un outillage spécifique et adapté aux besoins du bulletin, dans un espace nommé.

Penser en mode HTML, c'est avant tout placer en avant la sémantique du texte, son organisation… les différents rendus graphiques pouvant alors s'appuyer sur une base solide.

Depuis Latex

Côté auteur, au risque de déplaire aux *afficionados*, Latex, qui est largement adopté dans la communauté scientifique, à juste titre, reste un outil lourd (plusieurs gigas d'installation, des centaines de modules optionnels) et complexe. Il est remarquable dans ses facultés de composition (et inégalé en ce qui concerne la justification des paragraphes et la composition des mathématiques). Il incite cependant, par la complexité de son paramétrage, à se conformer à un modèle, tout en permettant bien des bidouilles locales qui font entrer beaucoup du travail de composition dans le processus d'écriture, ce qui ne devrait pas être le cas à ce stade[12].

Recevant tantôt du Word, tantôt du Latex, et ayant besoin d'unifier tout cela pour y travailler, nous ne souhaitions pas nous contenter d'un simple export. Nous avons choisi de faire d'abord converger le contenu de tous les tapuscrits en Asciidoc pour pouvoir continuer à les travailler, les mettre en forme puis échanger des épreuves avec les auteurs, autour d'un format unifié de travail et de conservation avant de le publier sous la forme HTML avec quelques transformations quand c'est nécessaire.

Nous avons exploité également la capacité d'Asciidoc à conserver telles quelles les commandes Latex pour avoir un meilleur rendu des équations (mode STEM), ainsi que Bibtex pour les bibliographies fournies dans ce format. Nous avons donc développé deux convertisseurs dédiés qui convergent depuis les deux formats sources et préparent un dossier de travail extrayant au besoin les images incluses dans le document Word, en ajoutant un fichier de métadonnées pour les légendes, crédits, licences et alternatives accessibles à ces images.

On ne l'imagine pas toujours, mais c'est souvent là que le travail commence vraiment (la fameuse *émendation* des éditeurs depuis la Renaissance) : préparation, relecture, discussion, complétion, harmonisation, report des corrections d'auteur, etc. Le tout autour d'épreuves composées et unifiées. Pour cela on passe le relais à HTML et CSS au moyen du générateur de site statique Eleventy (ou 11ty)[13]. Nous avons également profité de l'occasion de cette refonte pour proposer quelques recommandations aux auteurs de *1024*, qui, espérons-le, en feront bon usage. Certains essaieront peut-être même d'envoyer leur contribution directement en format texte, Markdown ou Asciidoc !

La correction et un Git

Le rôle d'une chaîne de publication ne se limite pas à l'organisation du texte, mais se doit d'organiser également les échanges entre les auteurs et les éditeurs. Avant de publier, il faut corriger (ou *émender*, terme un peu moins

12. Daniel Allington, *The Latex fetish (Or: Don't write in Latex! It's just for typesetting)*, 2016. `http://www.danielallington.net/2016/09/the-latex-fetish/`.
13. `https://www.11ty.dev/`.

sévère que *correction*, pour désigner un travail éditorial qui construit une complicité souvent fructueuse entre l'éditeur, le correcteur et l'auteur). C'est ce jeu d'allers-retours et de regards collectifs dans la fabrication du texte (qui deviendra ensuite livre ou revue) qui fait toute la richesse du processus éditorial et bien souvent toute la différence avec l'auto-édition.

Nous avons ainsi ajouté une feuille de style adaptée au flux de corrections qui permet d'annoter et de réviser de manière visible et réversible le texte. Les annotations portées par l'éditeur apparaissent en couleur dans le texte qui est retourné aux auteurs en format pdf ; les commentaires sont possibles. Les échanges peuvent s'organiser autour du texte, de manière intégrée, jusqu'à validation de la version définitive. Nous espérons que cela enrichira les échanges entre la rédaction et la communauté des auteurs.

L'avantage de travailler avec des formats ouverts et textuels, est qu'on peut disposer des dépôts de code et de leurs fonctionnalités pour partager, suivre, collaborer et conserver tous les états et toutes les versions du travail en cours. Nous nous appuyons sur Git et Gitlab pour cela. Toute l'équipe peut être synchrone, un éditeur en ligne associé à Gitlab permet même de participer au projet sans logiciel dédié (à l'exception d'un navigateur Web, bien entendu).

L'intégration avec Eleventy/11ty

C'est à l'occasion de la conception des publications du Louvre que nous avons pratiqué 11ty, un générateur de site statique écrit en Javascript, qui permet d'effectuer les conversions et transformations du format source, à travers des templates, vers des différents dossiers de publication (en ligne ou pour l'impression). Les pages Web destinées à l'impression sont prévisualisées dans le navigateur Web, grâce à PagedJS[14], *polyfill* (ou autre logiciel supplétif), qui permet de supporter la recommandation CSS pour les médias paginés du W3C, une recommandation qui n'est pas (encore) implémentée en standard dans les navigateurs. Julien Taquet, qui a longtemps travaillé sur le développement de pagedJS, ainsi que sur 11ty, nous a rejoint pour résoudre quelques casse-tête.

Grâce à cette structure, l'image de la double page imprimée est présentée dans le navigateur entièrement en HTML, sans avoir à installer d'autre logiciel, et permet d'utiliser les outils d'inspection pour évaluer les problèmes qui peuvent être rencontrés. La composition des équations Latex dans le navigateur est prise en charge par MathJax. Le fichier PDF pour l'imprimeur est construit au moyen de la commande d'impression du navigateur appliquée au document en cours.

Chaque article se voit attribuer des métadonnées et le DOI qui permet de déclarer son emplacement. On peut ensuite déposer le dossier Web du numéro complet obtenu via 11ty sur le serveur Web de la revue, simplement par FTP. Et voilà !

14. https://pagedjs.org.

La version Web intégrale

Cette méthode apporte un bénéfice non négligeable : à partir du code HTML balisant le contenu, il est possible de développer une autre feuille de style pour rendre le fichier comme une page Web. La version Web intégrale des articles de la revue peut ainsi, modulo les ajustements nécessaires (il s'agit de deux éditions différentes d'une même source, donc deux adaptations à des supports de lecture différents) devenir lisible en ligne. Et ceci sur n'importe quel terminal, notamment des téléphones mobiles ou des tablettes, grâce à une feuille de style *responsive*. C'est un progrès pour la revue et son ouverture. Cette page permettrait également le téléchargement de formats détachés comme le PDF.

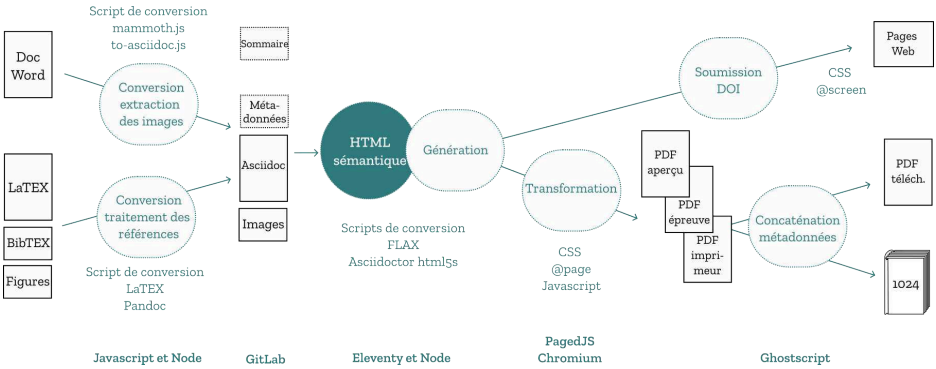


Fig. 3. Chaîne de traitement de *1024* : les moteurs.

Un modèle qui permet à chacun de se consacrer à son travail

On passe ici des étapes, mais nous avons présenté le déroulement général des opérations. Le modèle utilisé ne prétend pas être parfait, mais à chaque étape nous sommes libres d'améliorer et de modifier, et nous ne sommes nullement prisonniers d'un outil unique ou membres d'une chapelle. De plus, les outils du Web évoluant très rapidement, nous pouvons espérer profiter des progrès réalisés pour améliorer différents points de la chaîne de production proposée. Notre collaboration avec *1024* a été orientée vers l'outillage de chaque membre participant à l'élaboration de la revue. En synchronisant tout le monde autour d'une chaîne, nous espérons permettre à chacun de mieux se consacrer à son rôle : auteur, rédac'chef, correcteur, designer, compositeur, imprimeur, webmaster… et tout au bout les lecteurs, tout le monde doit y gagner. C'est ça une revue : un partage d'intelligence, pas artificielle.

Composer avec CSS

La mise en page avec HTML et CSS repose sur une recommandation du W3C pour les médias paginés qui n'a malheureusement pas été complètement implémentée par les navigateurs Web. Il existe plusieurs logiciels (propriétaires ou libres) qui comblent cette lacune. Nous avons choisi PagedJS qui agit dans un navigateur Web, est écrit en Javascript, et respecte intégralement cette recommandation et permet de s'appuyer sur la vitesse de traitement du navigateur Web pour obtenir un aperçu remis à jour en permanence. PagedJS est en outre extensible, ce qui permet de l'augmenter avec de petites améliorations personnelles que nous partageons volontiers avec la communauté des utilisateurs[15].

Les avantages et les limites de la mise en page Web

Mettre en page avec HTML et CSS présente des avantages très nets : vous l'avez compris, cela permet de séparer la structure du document de son apparence. D'un côté le document est décrit assez complètement (l'emploi de classes permettant d'étendre la sémantique comme détaillé plus haut), de l'autre CSS, est un très puissant moyen de spécifier l'aspect visuel de ces éléments, du général au particulier, de la structure de la page à la microtypographie. Le code HTML, CSS ou Javascript est lisible, et permet de travailler avec une grande variété d'outils, ce qui est important dans une chaîne ouverte ayant plusieurs intervenants. Quand on fait attention à respecter les emboîtages HTML et les règles de cascade des CSS, le code reste léger et clair, interprétable par des non-spécialistes. Enfin, ce code balisé et appuyé sur des normes solides et reconnues est suffisamment pérenne pour être conservé ou converti ultérieurement.

Bien entendu, la composition de texte avec ces outils n'atteint pas (encore) le niveau de finesse de rendu et d'intelligence de l'algorithme de Knuth et Plass[16]. Ni même du logiciel propriétaire standard de l'industrie Adobe InDesign. Mais les navigateurs — qui constituant un support de lecture très important désormais — progressent assez vite et pourraient rattraper ces autres logiciels assez rapidement. Pour aller plus loin encore dans le contrôle du mode paginé, nous avons progressé depuis 2018 et produit des extensions Javascript spécialisées. C'est un pari, avec des compromis. Il y a des limites, mais aussi des possibilités parfois plus avancées que dans les logiciels de PAO (par exemple bénéficier de la cascade et donc du contexte dans les spécifications de style, ou ne pas couper inopinément une ligne entre un article et le nom qui le suit).

CSS aide vraiment à prendre en charge différents formats, mais cela aussi avec des limites. Et on s'aperçoit à l'expérience que le multiformat implique

15. Lire à ce sujet la thèse de Julie Blanc, *Composer avec les technologies du Web, genèses instrumentales collectives pour le développement d'une communauté de pratique de designers graphiques*, 2023, université Paris 8 – ED 224 / EnsadLab / EUR ArTeC. En ligne : `http://phd.julie-blanc.fr/`.

16. Donald E. Knuth et Michael F. Plass, "Breaking paragraphs into lines", *Software-Practice and Experience*, Vol. 11, 1119-1184, 1981, `http://www.eprg.org/G53DOC/pdfs/knuth-plass-breaking.pdf`.

souvent des différences au niveau même du contenu (par exemple une vidéo ou une carte zoomable en ligne, respectivement remplacées par une image fixe et une échelle déterminée dans la version imprimée). Ce qui est plus complexe à gérer, comme nous l'avons écrit ailleurs[17]. Il faut intervenir sur la structure du document, ce qui pourrait être vu comme une rupture avec la notion de source unique. En pratique, on propose dans le même document HTML divers éléments orientés pour chaque usage. Dans la réalisation pour un usage donné, les éléments insérés pour d'autres usages sont rendus invisibles. Il s'agit là, nous en sommes bien conscients, d'un artifice qui permet de concilier la sauvegarde complète d'une source encodée, tout en s'adaptant aux divers médias de lecture.

Les défis

Notre expérience de HTML2Print a commencé il y a quelques années. Dans le premier ouvrage réalisé avec cette méthode, *Addiction sur ordonnance*, nous avons dans le colophon pointé les limites, soulevé le caractère expérimental… et invité les lecteurs à suivre la collection *Interventions* pour en évaluer les améliorations. Nos premiers ouvrages ne permettaient pas les notes de bas de page, ni la gestion des figures flottantes (le repositionnement arbitraire des figures au sein de la double page). Désormais nous avons intégré ces possibilités. Chaque réalisation apporte une meilleure maîtrise de la chaîne de production et la mise en place de subtilités qui nous aident à approcher la qualité typographique des autres outils. Dans cet ordre d'idée, réaliser une chaîne pour *1024*, comme auparavant pour le Louvre, nous incite à relever tout un lot de défis. Notamment la gestion du code et des équations, bête noire des logiciels de PAO et point fort de Tex.

On pourrait aborder d'autres aspects notables particuliers de ce chantier. Éditer une revue ou un bulletin est plus difficile qu'éditer une monographie. La diversité des textes, des types d'objet hors texte nous amène à nous adapter au fil des contributions et l'ensemble ne manque pas de surprises. Patrice Naudin, l'œil scrupuleux qui relit et amende les contributions jusqu'au «*bon à tirer*» pourrait en témoigner. Nous lui rendons ici hommage.

Faire de grands choix mais aussi les mille petits trucs et astuces qui permettent de caler une mise en page pour assurer un bon enchaînement des pages est quelque chose qui repose encore sur l'intelligence humaine. Cette humanité que ceux qui veulent tout automatiser essaieront de remplacer par une intelligence artificielle. Un jour probablement, mais pour le moment nous tablons, en ce qui nous concerne, non pas sur l'automatisation, mais sur l'outillage le plus correct possible des acteurs et de leurs compétences et le développement de modes de collaboration efficaces et agréables pour une équipe.

17. Nicolas Taffin, *op. cit.*

Un des aspects intéressants de tels développements est de pouvoir les partager, dans leur globalité (comme une chaîne) mais aussi à la découpe, un module pouvant à lui seul rendre bien des services. C'est Nous l'avons donc prévu, dès que nous aurons suffisamment « essuyé les plâtres », nous imaginons après un ou deux numéros. La documentation est faite au fil de l'eau dans cet esprit. Nous avons déjà en tête des idées d'évolutions futures et vous, lecteur, en aurez peut-être aussi d'autres à nous souffler.

Le Collège des sociétés savantes académiques de France

Amplifier la voix de la communauté académique

Patrick Lemaire[1]

Les scientifiques de tous bords semblent partager un mélange d'incompréhension et de désespoir face au décalage entre les travaux et connaissances scientifiques d'une part et les décisions politiques d'autre part, dans des domaines aussi divers que l'agriculture et l'alimentation, les transitions environnementale et énergétique, l'éducation scolaire ou la politique migratoire, pour ne citer que quelques exemples récents.

Pourquoi un tel décalage ? Comment œuvrer pour que le travail de millions de scientifiques de par le monde (dont environ 200000 en France) soit mieux pris en compte par les responsables politiques, pour le bénéfice des citoyens, et dans le respect de la fonction des élus ? Comment lutter contre la croissance préoccupante de la désinformation ? En bref, comment redéfinir la place des scientifiques dans la société et la place de la société dans le travail des scientifiques ?

Ces interrogations sont à la source de la création du Collège des Sociétés Savantes Académiques de France en février 2021, après 3 ans de gestation. La SIF, présente dès les premiers échanges, fait partie de ses membres fondateurs. Le Collège est actuellement un forum unique de par la diversité des disciplines académiques qui y sont représentées. On peut trouver de nombreuses informations sur le Collège et ses actions sur son site web.[2]

1. Biologiste et président du Collège des sociétés savantes académiques de France.
2. https://societes-savantes.fr.

DOI:10.48556/SIF.1024.24.73

Ce texte présente les missions et les valeurs du Collège, puis, à l'aide de trois exemples, il illustre la force (et les limitations) d'une approche interdisciplinaire et collégiale. Il se termine par une réflexion personnelle sur une nécessaire extension des formes d'engagement collectif des scientifiques.

Les missions du Collège

Les crises environnementale, sociale et énergétique actuelles sont presque toutes de nature très interdisciplinaire, associant des composantes de sciences et technologies et de sciences humaines et sociales. De ce fait, chaque société savante disciplinaire ne peut dans son expression publique individuelle prendre position que sur une partie du problème. La mission principale du Collège est de mettre en réseau les sociétés savantes de l'ensemble des disciplines académiques afin de renforcer la voix du monde académique dans le débat public.

De plus, le périmètre d'action de la plupart des sociétés savantes est centré sur le monde académique, sans relai efficace de sa parole dans la société. Le Collège a donc dès le début prévu d'inclure en son sein des associations à l'interface avec la société française et qui, sans être des sociétés savantes, poursuivent avec celles-ci des buts communs. Il s'agit notamment d'associations de professeurs de l'Éducation nationale, de médiation scientifique, d'éducation populaire ou de promotion de l'inclusivité dans le monde académique. Vous pourrez retrouver sur le site web du Collège[3] les 67 sociétés savantes et 13 membres associés qui le composent. Les sociétés savantes membres du collège comptent collectivement près de 50000 adhérents.

Les valeurs du Collège

Pour peser dans le débat public le Collège affirme plusieurs valeurs : l'indépendance vis-à-vis des pouvoirs publics, le non-partisanisme politique, la collégialité, l'interdisciplinarité et l'inclusion.

La première de ces valeurs est l'indépendance vis-à-vis des pouvoirs publics et des groupes de pression. Nous avons pour cela rédigé une charte des sociétés savantes académiques[4] que les sociétés membres signent lors de leur entrée dans le Collège, et une charte des membres associés[5]. Ainsi, seules les associations ayant un statut d'association loi 1901 sont éligibles, ce qui distingue le Collège des structures régies ou fondées par décrets, comme les académies des sciences, des sciences morales, des technologies, de médecine, etc.

3. https://societes-savantes.fr/qui-sommes-nous/.

4. https://societes-savantes.fr/wp-content/uploads/2020/01/CHARTE-DES-SOCIETES-SAVANTES-ACADEMIQUES-DE-FRANCE-2019-02-03.pdf.

5. https://societes-savantes.fr/wp-content/uploads/2021/10/2021-10-18_CHARTE_membres-associes-du-College.pdf.

La seconde valeur est le non-partisanisme politique. Les actions du Collège sont par essence éminemment politiques, puisqu'elles ont pour but d'agir sur la société française. Mais elles doivent exclure toute position partisane. D'une part, parce qu'elles ont pour ambition d'être fondées sur des faits scientifiques et non sur des idéologies. D'autre part parce qu'un positionnement partisan contribuerait à polariser la parole des scientifiques et conduirait nécessairement à s'aliéner une partie des citoyens, réduisant ainsi l'impact des actions.

La troisième valeur est la collégialité. Le Collège représentant ses membres, il ne peut donc prendre d'initiative majeure sans en référer à ceux-ci. Ses prises de positions écrites sont le résultat d'un processus — parfois long — d'écriture collective qui s'est jusqu'à maintenant terminé par la signature de chacun des membres favorables au texte. Si la longueur du processus peut être vue comme un obstacle à une réaction rapide à l'actualité, il permet aussi d'échapper à une forme de tyrannie de l'actualité où la nécessité perçue de commenter chaque décision politique consomme l'énergie nécessaire à une réflexion de fond.

Quatrième valeur : l'interdisciplinarité. L'action publique est un processus complexe. Prenons l'exemple, brûlant, de la transition environnementale. Décarboner l'économie, transformer les pratiques agricoles pour assurer la durabilité de ce secteur actuellement en crise profonde, ne peuvent reposer sur des approches purement technologiques : il faut aussi repenser les modèles économiques, comprendre les ressorts des comportements individuels des acteurs, et — de manière croissante — lutter contre la désinformation. Pour cela, sciences et technologies, sciences humaines et sociales, doivent apprendre à travailler ensemble et à se respecter malgré des valeurs épistémiques distinctes. A la différence des académies, spécialisées, des universités, pluridisciplinaires mais où le dialogue académique entre composantes est rarement soutenu, le Collège aspire à développer le dialogue entre disciplines.

Enfin la cinquième valeur est l'inclusion, parce que la lutte contre les comportements discriminatoires, vis-à-vis des femmes notamment mais pas seulement, est une nécessité pour que le talent de chacun puisse s'exprimer.

Le positionnement du Collège par rapport à d'autres mouvements de scientifiques.

Des mouvements comme Sauvons la recherche[6], Sauvons l'université[7], Sciences en Marche[8] ou Rogue ESR[9] ont été initialement formés pour dénoncer les mauvaises conditions de travail et l'insuffisance du soutien financier et en ressources humaines apporté par l'état à la recherche publique et à l'enseignement supérieur.

6. https://fr.wikipedia.org/wiki/Sauvons_la_recherche.
7. http://www.sauvonsluniversite.com/spip.php?rubrique3.
8. http://sciencesenmarche.org/fr/.
9. https://rogueesr.fr/.

Leurs actions avaient ainsi pour but principal un renforcement de la politique des sciences et du service public de l'enseignement supérieur en France. Il s'agissait aussi souvent pour ces mouvements militants de promouvoir une organisation collégiale de l'ESR, en opposition avec les règles du New Public Management.

Le Collège, lui, a été créé avec l'objectif prioritaire d'améliorer la prise en compte des connaissances et des méthodes scientifiques dans l'action publique. Il s'agit donc d'améliorer l'action publique **par** les sciences, et non de renforcer l'action publique **pour** les sciences. Les deux approches sont néanmoins complémentaires : il est peu probable que l'on arrive à convaincre les décideurs politiques de renforcer leur investissement dans l'ensemble des champs scientifiques tant que leur vision de l'apport des sciences à la société française se limite principalement à la production d'innovations technologiques favorisant la croissance économique à court terme.

Quelques exemples d'actions ayant un eu un impact dans le débat public.

Le Collège mène de nombreuses actions, dans la limite des moyens financiers et des bras disponibles. Ces actions incluent l'organisation de journées de réflexion, d'un congrès annuel, la publication de tribunes et de notes de synthèse sur des sujets de société, l'organisation de sondages auprès de la communauté académique… Plutôt que de faire un panorama de l'ensemble de ces actions, voici trois exemples d'actions transversales de natures différentes qui tirent pleinement parti de la diversité thématique du Collège et qui ont eu un impact médiatique ou qui ouvrent la possibilité d'actions nouvelles.

Repenser la formation des enseignants du primaire et du secondaire.

La commission enseignement scolaire[10], coordonnée par Louise Nyssen, a été la première commission thématique du Collège. Elle concentre actuellement ses travaux sur les formations initiale et (prochainement) continue des enseignants, à un moment marqué par une forte crise des vocations. La commission enseignement conduit également une réflexion sur les interventions des chercheurs en milieu scolaire, dans les disciplines de sciences expérimentales. Cette réflexion a été initiée par l'annonce du plan « Sciences » en 2022 et se poursuit, entre autres, par l'organisation d'une journée d'étude le 12 novembre à Paris[11].

Pourquoi s'intéresser à ce thème ? D'une part, la formation initiale des enseignants se fait majoritairement au sein des universités. Le monde académique

10. https://societes-savantes.fr/commissions-thematiques-du-college-des-societes-savantes-academiques-de-france.

11. https://societes-savantes.fr/accompagner-les-enseignants-du-primaire-en-science-comment-faire/.

est donc légitime à se saisir de ce thème. De plus, un véritable mal-être enseignant se propage depuis de nombreuses années, qui aboutit actuellement à ce que l'Éducation nationale ne réussisse plus à recruter des enseignants titulaires en nombre suffisant. Ce manque d'attractivité a conduit à ce qu'au moins un enseignant manque dans la moitié des établissements à la rentrée 2024, malgré le recrutement de contractuels. Redonner leur attractivité et leur sens aux métiers de l'enseignement en formant mieux les enseignants, futurs ou en poste, aurait des conséquences importantes sur la formation des jeunes et donc sur la société française dans son ensemble.

La formation initiale des enseignants a « subi » quatre réformes depuis 2010, aucune ne cherchant à tirer les leçons des précédentes… Cette succession de réformes, préparées dans la hâte, épuise les enseignants et participe à la désaffection pour ce métier. La commission enseignement scolaire a publié à l'automne 2022 une note de synthèse intitulée « Repenser *collectivement* la formation initiale des enseignants »[12] qui dresse un bilan des principaux écueils auxquels se heurte la formation des enseignants en master « Métier de l'enseignement, de l'éducation et de la formation » (MEEF) : organisation difficile, entrée dans le métier de plus en plus compliquée, aggravation des inégalités sociales, moindre maîtrise des connaissances disciplinaires… tandis que les stages, pivot de la formation professionnelle, concentrent les problèmes.

Cette note de synthèse, assez technique, a été communiquée au cabinet du ministre de l'Éducation nationale et à la présidente de la Commission des affaires culturelles de l'Assemblée nationale de l'époque, Isabelle Rauch, qui ont tous deux reçu une délégation du Collège. Du fait de la valse des ministres de l'Éducation nationale et de leurs cabinets, les rendez-vous ont dû être répétés avec chaque nouveau cabinet. La commission a été cette année auditionnée sur le même sujet par la mission « Exigences des savoirs »[13] et par la Mission d'information de l'Assemblée nationale « sur le recrutement et la formation du personnel enseignant dans les collèges et lycées publics »[14]. Et puis l'Assemblée a été dissoute, gelant la réforme. Le « choc des savoirs » tant voulu par Gabriel Attal ne verra pas le jour, du moins pas en 2025…

Plus généralement, ces textes et actions illustrent plusieurs leçons qui découlent de notre expérience. La première est que l'obtention de rendez-vous ministériels ou parlementaires est grandement facilitée par la publication d'un texte de fond, assez technique, et par sa diffusion dans les ministères et parmi les parlementaires. La seconde est qu'il ne suffit pas de publier un texte pour être écouté, il faut en assurer la promotion et également le décliner sous forme

12. https://societes-savantes.fr/wp-content/uploads/2022/10/2022-10-13-Repenser-collectivement-la-formation-initiale-des-enseignants-final-.pdf.

13. https://www.education.gouv.fr/une-mission-exigence-des-savoirs-pour-elever-le-niveau-de-notre-ecole-379788.

14. https://www.assemblee-nationale.fr/dyn/16/organes/commissions-permanentes/affaires-culturelles/missions-de-la-commission/mi-personnel-enseignant.

de tribune grand public, dans Le Monde dans ce cas[15], afin d'attirer l'attention des journalistes sur le problème traité mais aussi sur le travail du Collège.

Enfin, et je reviendrai sur ce point dans la dernière partie de ce texte, être écouté signifie rarement être entendu : face à une difficulté, il est souvent politiquement plus rentable de conduire une réforme superficielle, même inefficace ou contreproductive, que de prendre le temps de construire un travail de fond. Dans le cas de la formation des enseignants, changer la place du concours dans leur cursus est plus facile à faire (et à expliquer à l'opinion) que de penser de manière intégrée les besoins en formation des enseignants…

« Femmes en tête » : changer la perception des femmes scientifiques pour encourager les jeunes filles vers ces carrières.

Si le rapport femmes / hommes varie en fonction des domaines, seuls 37,8 % des chargés de recherche (CR) au CNRS sont des femmes. Cette proportion baisse à 31,7 % pour les directeurs et directrices de recherche (DR) (et à 30 % pour les DR de classe exceptionnelle). La part des femmes évolue certes favorablement mais bien trop lentement : lors des concours CR et DR2 en 2024, les femmes représentaient respectivement 44 % et 35 % des candidats admis sur première liste, les pourcentages étant tirés vers le bas dans les disciplines de sciences « dures/exactes » et technologiques (36 % et 27 % de femmes dans les recrutés CR et DR, respectivement). De plus, les femmes restent minoritaires parmi les doctorants.

Attirer davantage de femmes vers les carrières de recherche implique de lutter contre les stéréotypes de genre au sein même du monde académique. A l'invitation de Femmes & Sciences, le Collège a ainsi signé la convention pour une communication sans stéréotype de sexe[16] du Haut conseil à l'égalité entre les femmes et les hommes.

Il faut aussi changer l'image des femmes scientifiques et de leurs carrières auprès des jeunes filles. Le Collège organise chaque année depuis 2022, à l'occasion de la « Journée internationale des droits des femmes », l'opération « Femmes en tête »[17], qui met en lumière une sélection collective d'une dizaine de femmes remarquables du monde de la recherche, issues de toutes les disciplines, de tous âges et de tous niveaux de carrière, qui font la diversité et la force de la recherche actuelle.

Les femmes mises en avant dans ce type d'opérations donnent parfois le sentiment d'un « effet Marie Curie » : pour réussir en sciences, une femme doit être exceptionnelle, et donc intimidante et inaccessible. L'optique de « Femmes en tête » est différente. D'une part, la sélection « Femmes en tête » est une sélection collective et non un prix individuel. Les 30 premières « Femmes en

15. https://societes-savantes.fr/publication-dans-le-monde-de-leducation-de-la-tribune-penurie-denseignants-faute-de-bilan-des-reformes-precedentes-lannonce-dun-deplacement-du-concou/.
16. https://societes-savantes.fr/signature-convention-communication-sans-stereotype-de-sexe/.
17. https://societes-savantes.fr/actions/femmes-en-tete/.

tête », parmi lesquelles 3 informaticiennes (Daphné Giorgi[18], Irène Marcovici[19] et Meghyn Bienvenu[20]), ont été collectivement choisies parce que leur profil en faisait des rôles modèles accessibles qui incarnent auprès des jeunes collégiennes et lycéennes les multiples formes d'épanouissement d'une carrière scientifique féminine.

Collectivement les « Femmes en tête » constituent de plus un groupe unique de par sa diversité de parcours, d'âges et de disciplines. Nous avons comme objectif de le constituer en réseau et de lui donner les moyens de conduire des réflexions et de faire des propositions sur (par exemple) l'interdisciplinarité, l'inclusivité ou l'intergénérationalité dans l'ESR.

Créer les conditions d'un éclairage scientifique des politiques publiques nationales et locales

La décision politique est un processus complexe qui intègre les interventions de multiples acteurs politiques, économiques, administratifs ou associatifs, dotés de moyens très inégaux pour accéder à la connaissance et pour se faire entendre. Plusieurs conseils scientifiques existent au niveau national (par exemple : Haut conseil pour le climat, comité national consultatif d'éthique, Haut conseil de la Santé publique) ou régional (conseils scientifiques régionaux). Des avis scientifiques de grande qualité sont également produits par les académies et l'Office parlementaire d'évaluation des choix scientifiques et techniques (OPECST). Cependant, la pénétration de ces avis dans le débat public souffre de leur émiettement entre une multiplicité d'acteurs mal identifiés et travaillant souvent sans concertation.

Dans une note de synthèse publiée le 30 septembre 2023[21], 31 sociétés et associations du Collège des sociétés savantes académiques de France, dont la SIF, ont fait le constat de la trop grande distance qui sépare les responsables politiques des sciences. Pour y remédier, le texte propose la création d'un dispositif de conseil scientifique gouvernemental, placé auprès du premier ministre, et qui aurait quatre missions principales.

1. S'assurer que les élus et ministres aient accès à des synthèses des données scientifiques pouvant éclairer leurs décisions et qu'ils soient à même de les comprendre et de les prendre en compte. Évaluer a posteriori la compatibilité des décisions prises avec les consensus scientifiques.
2. Développer la présence de scientifiques au sein des administrations, afin de faciliter la mise en œuvre opérationnelle de politiques publiques éclairées par les sciences.

18. https://societes-savantes.fr/femmes-en-tete-2024-daphne-giorgi/.
19. https://societes-savantes.fr/femmes-en-tete-2023-irene-marcovici/.
20. https://societes-savantes.fr/femmes-en-tete-2022-meghyn-bienvenu/.
21. https://societes-savantes.fr/publication-de-la-note-reflexions-sur-les-missions-dun-conseiller-scientifique-du-gouvernement/.

3. Renforcer la culture scientifique du public et des journalistes, incluant à la fois les démarches et les savoirs fondamentaux dans un large éventail de sciences technologiques, expérimentales, humaines et sociales.
4. Renforcer la présence et l'influence de la France dans les organismes internationaux de concertation sur les grands défis planétaires.

La note évoquait la possibilité que la personne à la tête de ce dispositif incarnerait et agirait comme porte-parole d'une autorité administrative indépendante (AAI)[22] dotée de plusieurs conseils scientifiques et qui pourrait regrouper certaines structures de conseil déjà existantes.

Ces propositions ont été envoyées avec demande de rendez-vous à des parlementaires, à la ministre de l'Enseignement supérieur et de la recherche, au cabinet du premier ministre de l'époque, et à celui du président de la République. Si les parlementaires sont restés muets, nous avons en revanche été reçus par la conseillère recherche du MESR, puis, ensemble, par les conseillers recherche de Matignon et de l'Élysée. Une des pierres d'achoppement de cette seconde discussion a été le positionnement de ce dispositif auprès du gouvernement, au plus près de l'action, ou auprès du président. L'Élysée tenait à un positionnement au plus près du président. La création peu de temps après du Conseil présidentiel de la science (avec de vagues missions de conseil stratégique de politique scientifique au bénéfice du seul président de la République et non d'éclairage scientifique des politiques publiques…) a clairement montré où les arbitrages sont faits…

Ces propositions nous ont aussi permis d'amorcer des discussions fécondes avec des P.-D.G. ou responsables de grands organismes (CNRS, IFREMER, IRD, France Universités) ainsi qu'avec deux anciens conseillers scientifiques en chef étrangers : Sir Patrick Vallance (RU pendant le Covid, maintenant ministre des sciences dans le gouvernement Starmer) et Peter Gluckman (premier conseiller scientifique en chef de la NZ et actuel président de l'« International Network for Government Scientific Advice — INGSA »[23]). Une synthèse de ces discussions est en cours pour affiner le document initial. Si les propositions du Collège n'ont pas (encore) fait bouger les lignes, elles ont néanmoins eu une bonne visibilité médiatique à travers une tribune[24], et un entretien[25] dans *Le Monde*, un débat avec Sylvie Retailleau sur France Culture[26], et enfin un éditorial dans *Science*[27]…

22. https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033897475.
23. https://ingsa.org/.
24. https://societes-savantes.fr/publication-de-la-tribune-renforcer-le-poids-des-connaissances-scientifiques-dans-la-decision-publique/.
25. https://societes-savantes.fr/reaction-a-chaud-du-president-du-college-des-societes-savantes-suite-aux-annonces-sur-la-recherche-demmanuel-macron/.
26. https://societes-savantes.fr/debat-radiophonique-entre-patrick-lemaire-president-du-college-et-sylvie-retailleau-ministre-de-lesr/.
27. https://societes-savantes.fr/editorial-du-19-avril-de-la-revue-science-france-needs-a-chief-science-adviser/.

Succès, échec et visibilité du Collège

Un premier bilan des actions du Collège montre à la fois le chemin parcouru et celui qui reste à parcourir.

La première constatation est que le Collège devient peu à peu visible :

- dans la communauté académique : le nombre de sociétés savantes membres est passé de 44 à 67 en 3 ans, et une première société de droit nous a rejoint ;
- dans le monde associatif : le nombre de membres associés est passé de 5 à 13 ;
- dans la presse nationale (Le Monde, Libération, l'Humanité) et régionale ;
- des pouvoirs exécutif (MESR, MENJS, Matignon, Élysée) et législatif (Commission de la culture de l'Assemblée nationale) ;
- à l'international : rencontre de conseillers scientifique étrangers, tribune dans Science, commentaires sur la politique scientifique française dans Science et nature, demande d'adhésion prochaine à l'INGSA.

Cette visibilité résulte de plusieurs choix. Le premier est de ne pas se disperser, en évitant notamment de réagir à chaud à l'actualité. Nos procédures de réaction (communiqué coécrits avec les membres) sont suffisamment longues pour nous dissuader de réagir à chaud, sauf dans des cas suffisamment graves pour que CA réagisse en son nom propre (voir les communiqués sur la loi immigration de décembre 2023[28] et sur les élections législatives de juin 2024[29]). Le second est de chercher à redevenir maîtres du calendrier, en identifiant à l'avance des thèmes qui vont faire l'actualité : ainsi de la formation des enseignants, ou de l'éclairage scientifique des décisions politiques. Le troisième enfin est de s'appuyer sur le travail de fond de rédaction de textes de synthèse pour solliciter des rendez-vous et rédiger des tribunes.

Paradoxalement, cette visibilité médiatique ne se traduit pas par une large connaissance du Collège au sein de la communauté académique et les adhérents du nombre croissant de sociétés membres : les comptes réseaux sociaux du Collège ont moins de 500 abonnés, la page du bulletin d'information du Collège, Collège Info[30], est consulté moins de 500 fois à chaque nouveau numéro. Pour accroître notre visibilité, nous avons commencé à constituer un réseau des responsables communication des membres du Collège afin de faire vivre davantage nos réseaux sociaux. Si vous aimez communiquer, rejoignez l'équipe communication[31] !

28. https://societes-savantes.fr/la-loi-immigration-xenophobe-et-populiste-ignore-sciemment-les-travaux-scientifiques/.
29. https://societes-savantes.fr/communique-elections-legislatives-2024-1/.
30. https://societes-savantes.fr/bulletin-college-info/.
31. communication@societes-savantes.fr.

Au-delà, nous avons créé cette année la première édition du concours photo « Sciences grand angle »[32] pour augmenter la visibilité du Collège auprès des laboratoires et nous avons décidé de relancer prochainement une grande consultation sur l'état de la recherche, de l'enseignement supérieur et des relations entre science et société, à l'image de ce que nous avions fait lors de la rédaction de la Loi de programmation de la recherche[33].

Enfin, pour se rapprocher des laboratoires (et des responsables politiques locaux) il faut mailler le territoire. Le Collège des Sociétés savantes encourage ainsi les adhérents des sociétés savantes sur un même site à se rapprocher pour créer des antennes qui, en plus de relayer les actions nationales du Collège, conduiront des actions locales auprès des institutions, élus ou collectivités locales, monteront des événements transdisciplinaires, ou signeront des textes dans la presse locale. Parce qu'elles rapprochent le Collège des scientifiques, les antennes sont aussi de potentielles pépinières de futurs responsables nationaux du Collège. Une première antenne[34], coordonnée par Florence March (SHS, Société d'études anglo-américaines des XVII^e^ et XVIII^e^ siècles) et Gwladys Toulemonde (Sciences et technologies, Société française de statistique), a été crée à Montpellier en février 2023.

Chaque société membre peut contribuer à la vie du Collège et à l'augmentation de sa visibilité au sein de la communauté : en présentant chaque année certains de ses adhérents aux élections de renouvellement partiel du CA, à l'opération « Femmes en tête », aux commissions thématiques, en encourageant les soumissions au concours photo « Sciences grand angle », en participant à la rédaction des textes et tribunes, en étant représentée aux congrès annuels… ou en ouvrant son bulletin 1024 à la présentation du Collège que vous êtes en train de lire. Mais nous constatons aussi avec plaisir aussi que les sociétés membres sollicitent de manière croissante le soutien du Collège et de ses membres pour certaines des actions qu'elles aimeraient porter. L'union fait la force.

Accepter de sortir de notre zone de confort

La visibilité médiatique et auprès des responsables politiques du Collège ne suffit pas encore pour avoir des effets directs sur les politiques publiques. Comment amplifier l'impact des actions du Collège et plus généralement de la communauté scientifique ? Le prochain congrès du Collège, organisé par l'antenne locale de Montpellier les 3 et 4 février 2025, aura justement pour thème : « Sciences, responsabilité et engagement : quelles stratégies pour compter dans l'action publique ? »[35]. Pour la première fois, la journée de réflexion du congrès

32. https://societes-savantes.fr/concours-photo-sciences-grand-angle/sga-2024/.
33. https://societes-savantes.fr/dossiers/loi-de-programmation-pour-la-recherche/sondages/.
34. https://societes-savantes.fr/antennes-locales/antenne-montpellier/.
35. https://societes-savantes.fr/actions/congres-du-college-des-societes-savantes-academiques-de-france/congres-2025-montpellier/.

du Collège sera largement ouverte à la communauté académique. Notez-la dès maintenant dans vos agendas !

A titre personnel, je pense que pour avoir un réel impact dans la société la communauté académique doit sortir d'une zone de confort purement académique et souvent mono-disciplinaire.

Ainsi, pour formuler des recommandations sur des sujets de société complexes pouvant aller jusqu'à une remise en question des fondements de nos sociétés, il nous faut sortir d'une conception trop étroite — et parfois fantasmée — de la neutralité axiologique de Max Weber qui pousse à séparer les jugements basés sur les seuls faits et ceux basés sur des valeurs. De même qu'au siècle des Lumières, la promotion d'approches empiriques et rationnelles allait de pair avec un message d'émancipation intellectuelle et sociale, nous ne devons pas hésiter à mettre les approches scientifiques au service de la construction d'un modèle de société compatible avec la préservation de la planète. Pour cela, il faut à la fois développer des innovations technologiques et comprendre les moteurs des comportements individuels et collectifs, de consommation notamment. Or l'interdisciplinarité large associant sciences expérimentales et technologiques et sciences humaines et sociales reste le fait d'une minorité de projets scientifiques.

Cette interdisciplinarité gagnerait à s'étendre aux disciplines du droit, par exemple pour défendre la liberté académique contre les procédures judiciaires cherchant à brider la recherche sur des sujets sensibles. Ainsi la fondation Lejeune intente de nombreux procès pour annuler des autorisations de recherche, sur les embryons humains notamment[36]. Mais les voies judiciaires peuvent, inversement, aussi être utilisées en soutien à la prise en compte des travaux scientifiques par les gouvernements à l'image de « l'Affaire du siècle »[37] dont l'action a permis de condamner l'État pour inaction climatique.

36. https://www.lemonde.fr/idees/article/2017/03/30/nous-medecins-et-chercheurs-mettons-en-garde-contre-la-fondation-jerome-lejeune_5102969_3232.html.
37. https://laffairedusiecle.net/qui-sommes-nous/.

Pour une citoyenneté informatique

Plaidoyer pour une investigation en commun des codes sources publics

Marie Alauzen[1]

Une fois que vous avez accroché votre casque de vélo derrière la porte et posé votre tasse de café sur le bureau, vous vous installez, tout en saluant les collègues qui passent dans le couloir. Vous allumez votre machine et ouvrez la messagerie. Au milieu de la pile, bonne nouvelle, votre dernier manuscrit a été accepté avec des évaluations enthousiastes. Vous partirez donc dans quelques mois le présenter dans une conférence. Vous avez besoin d'une prise en charge de l'inscription, de l'hébergement, des billets de train et devez anticiper les dépenses courantes sur place. Ainsi, vous commencez votre journée en saisissant une demande dans l'application, qui notifiera le gestionnaire, puis la directrice du laboratoire de votre future mission.

Les cadres juridico-informatiques de l'expérience

Cette situation — qui pourrait sembler anodine et sans intérêt pour le *Bulletin de la Société d'informatique de France* — permet de mettre en lumière deux cadres juridico-informatiques qui méritent toute notre attention[2].

1. Université Paris-Dauphine PSL, CNRS LAMSADE UMR 7243.

2. La notion de cadres (*frames*) renvoie aux travaux fondateurs du sociologue Erving Goffman [6], qui nous invite à analyser la structure de l'expérience individuelle de la vie sociale en considérant des moments d'activité (comme s'installer à son poste de travail) qui s'inscrivent dans une écologie particulière où le langage corporel est indissociable des ressources mobilisables dans l'espace où se déroule cette activité (ici, à la fois la chaise, le bureau, l'ordinateur, la tasse de café et les applications).

DOI:10.48556/SIF.1024.24.85

Premièrement, l'installation à votre bureau et la présence de vos collègues est encadrée par une durée annuelle du travail fixée par l'article L611-1 du code de la fonction publique[3], un cycle hebdomadaire, des jours d'ouverture du site et des congés qui se déclarent dans une application décomptant vos 46 jours annuels. Une cascade d'articles des codes de la fonction publique et de la recherche, ainsi que des arrêtés, des décrets et des décisions de l'établissement règlementent donc l'organisation du laboratoire. Et ce, même s'ils ne contraignent pas fermement votre expérience, puisque votre badge fonctionne pendant les vacances et que nul ne viendra s'enquérir de l'heure nocturne à laquelle vous avez soumis ledit manuscrit.

Deuxièmement, le décret n°2006-781 du 3 juillet 2006 modifié par le décret n°2024-746 du 6 juillet 2024[4] (célébré par les syndicats à l'été 2024 !) fixe les conditions et les modalités de règlement des frais de déplacement des chercheurs en conférence. L'acceptation de votre communication n'est donc pas seulement un heureux événement épistémique, elle suppose aussi une forte composante matérielle : l'organisation d'une mission pour laquelle vous pouvez prétendre à une prise en charge du transport et à des indemnités.

Ici, en dépliant les cadres de cette situation, il apparaît que les lois et les règlements qui s'appliquent à tous les agents publics ont été transposés localement par les instituts de recherche et les établissements dans des instructions. Ces instructions ont ensuites été traduites par les directions des systèmes d'information (DSI) en des spécifications fonctionnelles à destination des prestataires ou des agents éditant les fameuses applications de réservation de mission et de gestion des absences. Puis, des agents de la DSI ont ensuite conçu et mis à jour des guides pour leurs utilisateurs (chercheurs, gestionnaires, directeurs d'unité, agents comptables). Nous nommerons *altération informatique du droit* ce phénomène de transport de la référence juridique vers son opérationnalisation informatique — un phénomène à la fois très ordinaire et tellement ennuyeux[5] qu'il en est invisible (sinon très partiellement visible) [1].

L'altération informatique désigne le processus indissociablement social et technique au cours duquel des énoncés de droit (comme le deuxième alinéa de l'article L611-1 du code de la fonction publique[6] : « Le décompte du temps de travail est réalisé sur la base d'une durée annuelle de travail effectif de 1 607 heures ») sont déclinés dans des textes plus spécifiques, puis traduits dans des

3. https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000044425088/2022-03-01.

4. https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000049894058.

5. La qualification d'ennuyeux renvoie à l'expression malicieuse de la sociologue Susan Leigh Star, fondatrice de la Society of People Interested in Boring Things. Poursuivant le projet d'Anselm Strauss d' « étudier l'in-étudié », elle revendiquait l'examen des choses rébarbatives et triviales, a priori pensées comme étant indignes d'une étude scientifique. L'annuaire, les spécifications techniques des logiciels, les classifications standardisées des scientifiques ou encore des listes constituent, selon elle, des moyens privilégiés d'accéder aux décisions prises en coulisses et d'analyser la structuration des connaissances dans les infrastructures sur lesquelles repose notre société [5,12,20,21,22].

6. https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000044425088/2022-03-01.

programmes informatiques, afin de prendre effet. Il s'agit là de changements d'état successifs indispensables à la performativité[7] du droit. Pour autant, si cette altération est incontournable pour que les mots du droit déclenchent des actes, il reste à savoir qui réalise, contrôle, corrige cette chaîne d'écriture et de lecture et, surtout, comment ces acteurs opèrent cet irremplaçable travail d'interprétation du droit. En d'autres termes, l'énigme de cet article porte sur les *implications politiques* de l'altération informatique.

Afin d'intéresser la communauté de recherche en informatique à la résolution de l'énigme, nous localiserons d'abord le cœur du processus. Elle se produit dans l'écriture, la mise à jour et la maintenance des codes sources des infrastructures de service public situés dans les applications de gestion des congés et des missions, de calcul de la paie ou des impôts. Or, ces codes sources sont, par défaut, des documents publics pouvant être communiqués. Nous caractériserons ensuite trois usages qui ont été faits du premier code source rendu public en France, en déduisant l'exercice de la citoyenneté qui se dégage de ces usages. Enfin, j'inviterai la communauté à jouer un rôle moteur dans l'examen de l'altération informatique du droit, c'est-à-dire à mettre des savoirs professionnels à profit pour exercer une citoyenneté informatique.

Le droit de demander des comptes à l'administration

L'article 15 de la Déclaration des droits de l'homme et du citoyen de 1789 prévoyait que « la société a le droit de demander compte à tout agent public de son administration ». Toutefois, la *culture politique de la généralité* [18], qui a émergé à partir de la Révolution française autour des idées d'un face-à-face sans médiations entre l'individu et l'État, d'une démocratie immédiate et sans intermédiaires et d'une centralité de la loi, excluant l'existence d'interprétations par l'administration, dessine une tout autre relation entre les administrés et les administrations que celle dans laquelle ces dernières rendent compte de leurs actions. Comme l'historien Frédéric Graber l'a montré pour le cas des enquêtes administratives [7], la culture politique de la France qui s'est stabilisée au XIX^e^ siècle et déployée jusqu'à la fin du XX^e^ siècle est plutôt celle qui a nommé consentement l'acte d'un individu face aux décisions des autorités qui s'imposent à lui.

7. La performativité est un concept qui a connu une riche carrière à travers les sciences sociales. Je l'emploie ici pour souligner l'articulation entre un énoncé et l'action sociale qu'il accomplit sans masquer ni l'agencement sociotechnique nécessaire pour passer de l'un à l'autre ni le fait que l'énoncé (ici de droit) peut toujours échouer. Ainsi, un acte de langage juridique comme « Je vous marie » n'accomplit l'action sociale du mariage civil que s'il est pris dans un agencement sociotechnique pertinent : outre la légalité du mandat de l'édile et l'énonciation dans la salle des mariages de la mairie, les conditions de félicité de cet énoncé supposent un corpus de droit de la famille et la présence d'un registre pour le présent et le futur ouvrant des droits spécifiques aux époux en matière de fiscalité et de patrimoine. Il reste que performativité ne signifie ni performance réussie ni efficacité, car même dans cet agencement, l'énoncé du mariage reste vulnérable à l'échec. Il court le risque de se défaire ou de se transformer à chacune des épreuves (par exemple, si les témoins ne signaient pas l'acte au bon emplacement, si l'acte se perdait, etc.).

Consentement à l'impôt, consentement à un projet d'aménagement urbain, à l'édification d'une infrastructure ou à l'autorisation d'une industrie ont en commun le silence des administrés et la construction d'un point de vue surplombant de l'administration. C'est ainsi que, maigre compte-rendu, l'action administrative s'est affichée dans l'espace urbain, pour une durée limitée, afin de permettre aux administrés de prendre connaissance de certains actes [8]. Le propos était plus d'affirmer que l'administration a effectivement appelé tout le monde à en prendre connaissance que de rendre des comptes.

À partir des années 1960, un certain nombre d'attentes sociales ont émergé vis-à-vis de l'administration, la constituant en problème public ; cette mise en cause a nourri une *demande de réforme de l'État*. Comme le raconte le politiste Philippe Bezes [3, p. 127–203], des attentes se sont formées à la fois dans la dénonciation de la « bureaucratie » — au sens d'un jugement critique comprenant pêle-mêle : la paperasserie qui caractérise les procédures, la lenteur et l'anonymat du traitement, mais aussi l'insensibilité aux spécificités des administrés — et, en même temps, le rejet d'un État français qui serait trop centralisé et oppresseur. Les aspirations des mouvements sociaux dans le sillage de mai 1968 ont ensuite rendu le diagnostic d'une « crise de l'État » plus aigu et la quête d'une légitimité renouvelée par la publicisation de l'action administrative plus évidente. C'est ainsi que s'est opérée une socialisation de l'État et plus particulièrement de son administration, auparavant perçue comme étant un simple exécutant technique du politique, n'évoluant que dans des conflits connus d'une petite élite. Des réformateurs externes à l'administration sont apparus (professeurs de droit, journalistes, syndicalistes, partis d'opposition) et se sont coalisés pour faire émerger d'autres forces et institutions en mesure d'exercer une pression sur l'administration. C'est une configuration de réforme que Philippe Bezes nomme le « réformiste des contre-pouvoirs ». Sur le plan législatif, cette demande de réforme de l'État s'est traduite par trois textes votés à la fin des années 1970 : la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés[8], la loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal[9] et la loi n° 79-18 du 3 janvier 1979 sur les archives[10]. Saisie d'un projet de loi qui tendait à améliorer les relations entre l'administration et le public, l'Assemblée nationale transforma le texte qui devint le titre I^{er} de la loi du 17 juillet 1978, fixant un cadre de communication des documents administratifs et donnant naissance à une autorité administrative indépendante : la Commission d'accès aux documents administratifs (CADA).

8. https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000886460.
9. https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000339241.
10. https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000322519?isSuggest=true.

Depuis, les codes sources utilisés dans le cadre de l'exercice de missions de service public sont des *documents administratifs communicables au public*, par une demande directe à l'administration concernée ou, en cas de refus, à la CADA. Plus précisément, la CADA et le tribunal administratif de Paris ont tous deux considéré, lorsqu'ils ont été saisis, en 2015 et 2016, que les programmes informatiques, dans chaque version de leur code source, constituaient un document administratif communicable au public au sens de la loi de 1978[11]. Pour éviter que d'autres administrations ne contestent l'interprétation de la CADA et du tribunal administratif de Paris (soit de fait en ne fournissant pas le code, soit devant un tribunal), et pour appuyer par un texte de loi le mouvement d'ouverture des données et des algorithmes publics entrepris par la mission Étalab, le gouvernement a inscrit les codes sources[12] dans la liste des documents considérés comme étant publics par l'intermédiaire de la loi n° 2016-1321 du 7 octobre 2016 pour une République numérique[13]. C'est ainsi que l'article L300-2 du code des relations entre le public et l'administration[14] s'est étoffé.

Version antérieure à la loi pour une République numérique	**Version modifiée par la loi pour une République numérique**
Sont considérés comme documents administratifs, au sens des titres I^{er}, III et IV du présent livre, quels que soient leur date, leur lieu de conservation, leur forme et leur support, les documents produits ou reçus, dans le cadre de leur mission de service public, par l'État, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d'une telle mission. Constituent de tels documents notamment les dossiers, rapports, études, comptes rendus, procès-verbaux, statistiques, instructions, circulaires, notes et réponses ministérielles, correspondances, avis, prévisions et décisions.	Sont considérés comme documents administratifs, au sens des titres II^{e}, III et IV du présent livre, quels que soient leur date, leur lieu de conservation, leur forme et leur support, les documents produits ou reçus, dans le cadre de leur mission de service public, par l'État, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d'une telle mission. Constituent de tels documents notamment les dossiers, rapports, études, comptes rendus, procès-verbaux, statistiques, instructions, circulaires, notes et réponses ministérielles, correspondances, avis, prévisions, *codes sources* et décisions.

11. «Sont, en application des dispositions du titre I de la loi n° 78-753 du 17 juillet 1978, *au nombre des documents administratifs communicables, tous les programmes informatiques*. Peut ainsi être communiquée aux personnes en faisant la demande *chaque version du code source d'un même programme* informatique qui revêt le caractère de document administratif achevé sans que le caractère évolutif d'un programme informatique ne fasse obstacle au droit à communication de ce programme». Tribunal administratif de Paris, 5^{e} section, 2^{e} chambre, 10 mars 2016, n°1508951.

12. `https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000033202948`.

13. `https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746`.

14. `https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000033218936/2016-10-09`.

Même si la formule est mystérieuse et suscite des querelles d'interprétation parmi les professionnels du droit [9], il est indéniable que le droit de la société à demander des comptes à l'administration a progressé depuis son énonciation dans le tumulte de la Révolution. L'activité administrative est graduellement devenue l'objet d'une critique sociale, reconnue comme légitime et qui a amené, à partir des années 1970, l'institutionnalisation de contre-pouvoirs. Le droit d'accès aux documents administratifs, dont la CADA est garante, figure parmi ces derniers et s'applique aux codes sources des programmes utilisés dans le cadre des missions de service public. Ainsi, les codes des applications de gestion des congés et des missions évoquées en préambule sont, comme tant d'autres, communicables au public. Mais, que faire du code source de telles applications, qui font partie des cadres juridico-informatiques de l'expérience des chercheurs et, plus largement, de tous les citoyens ? Partant de l'ouverture du code source de l'impôt sur le revenu, nous passerons en revue trois attitudes correspondant à autant de modalités d'exercice de la citoyenneté, susceptibles d'inspirer des démarches vis-à-vis d'autres codes sources publics.

L'exercice de la citoyenneté à partir du droit d'accès

Le programme calculant l'impôt sur le revenu est intéressant à au moins trois égards. Historiquement, en France, il fut le premier code source d'une grande infrastructure administrative à être rendu public ; politiquement, il est l'un des fondements de la confiance au sein de la société moderne, matérialisant la fiction révolutionnaire du consentement à l'impôt et, techniquement, puisque tout l'impôt est défini par des règles inscrites dans la loi de finances et les textes d'application liés, il concentre le phénomène qui nous intéresse dans cet article : l'altération informatique du droit. En m'appuyant sur une enquête socio-historique sur le programme de calcul de l'impôt sur le revenu,[15] conçu à partir de 1987 à la direction générale des impôts et publié en 2015, je distinguerai trois exercices de la citoyenneté[16] : la citoyenneté comme témoignage, comme valorisation et comme vérification.

Commençons par la citoyenneté comme *témoignage*. Le 14 mai 2014, un étudiant de l'économiste Antoine Bozio à l'Institut des politiques publiques, collaborant à une modélisation informatique du système fiscal français nommée OpenFisca, a formulé lors d'un stage à Étalab une demande d'accès au code

15. Elle est le produit de la rencontre avec Denis Merigoux et Liane Huttner, en avril 2021, dont il sera question plus loin. Elle prolonge aussi mon enquête de thèse sur l'activité ordinaire de modernisation de l'État durant laquelle j'ai croisé plusieurs acteurs de cette histoire.

16. Je laisse volontairement de côté les interventions qui ont lieu à « l'intérieur de l'État », soit dans un processus qui relie l'administration fiscale à d'autres administrations — qu'il s'agisse de l'avis de la CNIL relatif au traitement des données personnelles, ou un audit de l'Agence nationale de la sécurité des systèmes d'information (ANSSI).

source du simulateur du site des impôts[17]. Devant le refus de la direction générale des finances publiques (DGFIP), le 19 novembre 2014, cet étudiant a saisi la CADA d'une demande de communication, afin de réutiliser le code de la simulation officielle dans le cadre de ses recherches et qu'il soit mis à la disposition d'Étalab sous licence libre.

Cette initiative relève de l'enquête : un petit groupe de personnes — un étudiant et son encadrant, l'Institut des politiques publiques, Étalab — a tenté de décortiquer la fiscalité française pour améliorer un objet technique (le simulateur OpenFisca) et pour produire de nouvelles connaissances (par exemple, sur les effets d'une réforme sur la redistribution de la richesse). Leur enquête supposait d'exercer une forme de soupçon sur le travail de la DGFIP, ne pouvant être levé que par un accès à la réalité : *voir le programme derrière l'interface en ligne*. Cette intervention microscopique, rendue possible par l'institutionnalisation de la transparence par les lois de 1978 et 1979 et la création de contre-pouvoirs, correspond à une façon de mettre en cause le caractère absolu de la puissance administrative. Elle peut s'apparenter à une épreuve au cours de laquelle le public mesure la force du droit dans et sur l'administration (notamment, lorsque le ministère public ne céda pas à la demande de la CADA et que l'étudiant dut saisir le juge administratif)[18]. Je propose de qualifier cette modalité d'exercice du droit d'accès la citoyenneté comme témoignage, car l'enquête consistait à voir le code du simulateur, puis répercuter ce qui avait été déduit du fonctionnement de l'administration fiscale dans ses activités (ici, en mettant à jour les paramètres du modèle de taxation de l'impôt sur le revenu)[19].

Venons ensuite à la citoyenneté comme *valorisation* de l'activité administrative. Il s'agit de la démarche entreprise par Étalab auprès de la DGFIP, d'abord sans succès, puis, à partir des discussions sur le projet de loi pour une République numérique à la fin de l'année 2015, contrainte par un arbitrage

17. Dans son analyse de la transformation de l'État par les communs numériques, le sociologue Sébastien Shutz retrace finement la mobilisation autour d'OpenFisca, qui va mener à cette demande d'accès. Elle est le fruit de la formulation d'un objectif commun entre, d'une part, Antoine Bozio, qui cherchait des financements pour l'Institut des politiques publiques qu'il dirigeait et ambitionnait d'alimenter le débat public en opposant, grâce à OpenFisca, des contre-calculs à l'administration, et, Henri Verdier, directeur d'Étalab, qui souhaitait légitimer l'action de « gouvernement ouvert » auprès des ministres [19].

18. La notion d'épreuve d'État, notamment mise au jour par le sociologue Dominique Linhardt, désigne une configuration historique dans laquelle la réalité de l'État se constitue comme un problème collectif et se spécifie dans des accomplissements pratiques. Ainsi, le terrorisme en Allemagne dans les années 1970, peut être compris comme une séquence historique au cours de laquelle la *Rote Armee Fraktion* affrontait l'État par des provocations en ville, au tribunal et en prison, qui furent autant de manières d'éprouver, dans les dispositifs mis en œuvre pour lutter contre cette guérilla urbaine, le caractère démocratique de la République fédérale (respect des droits humains en prison, caractère équitable du procès, statut des crimes politiques, etc.) [13].

19. J'écris volontairement « voir le code », car comme l'explique Sébastien Shutz : « Une fois obtenu, personne, ni à Etalab ni à l'IPP, n'a pu exploiter ce code, car il est écrit dans un langage informatique *ad hoc* et le coût de traduction pour l'utiliser dans Openfisca est trop élevé » [19, p. 858]. Nous verrons qu'il s'agit là d'une différence majeure avec la citoyenneté comme vérification, qui suppose des capacités à lire le code.

gouvernemental. Étalab incita la DGFIP à communiquer au sujet de la publication du code source de l'impôt sur le revenu, c'est-à-dire à faire de la future décision du juge, non pas un camouflet révélant l'opacité et étendant l'horizon du soupçon, mais une démarche pionnière, anticipant le projet de loi et prolongeant la configuration de réforme par les contre-pouvoirs précédemment évoquée [3]. L'idée, tributaire du contexte de réforme de l'État dont la mission Étalab, rattachée au Secrétariat général pour la modernisation de l'action publique (SGMAP), a été l'une des chevilles ouvrières, était de présenter l'ouverture de ce code comme une réalisation de « l'État plateforme ». L'État plateforme, qui constituait alors le principal mot d'ordre modernisateur [2], désignait un projet de redéfinition des relations administratives, supposant que les administrations mettent à disposition sur un site internet les données et les programmes qu'elles utilisent dans des formats ouverts, afin que des particuliers, des entreprises ou d'autres administrations puissent utiliser ces ressources pour créer de nouveaux services numériques, marchands ou non. Suivant cette logique de création de valeur, Étalab et la DGFIP organisèrent, les 1er et 2 avril 2016, le hackathon #CodeImpot, dans les locaux parisiens de la fondation Mozilla. Le code source, découvert sous le nom du langage de programmation inventé par l'administration fiscale à la fin des années 1980 pour les besoins du calcul, M, ainsi qu'un compilateur de M vers Python écrit par Étalab furent mis en ligne, afin que de petits groupes de bénévoles (principalement des agents publics ou des prestataires) initient une *exploration des chapitres du code et fassent émerger des entités temporaires, composites, souvent de très petite taille* (un outil pour référencer le nom des variables, un autre pour optimiser les déclarations communes des foyers ou maintenir OpenFisca). Les deux journées du hackathon organisées dans l'antenne de cette fondation californienne s'apparentaient moins à une enquête collective sur la vie administrative ou le droit fiscal, qu'ils ne furent caractéristiques de ce que la sociologue Lilly Irani nomme la « citoyenneté entrepreneuriale » [11]. En effet, il ne s'agissait pas seulement de produire des technologies, mais de célébrer un type de changement social prenant la Silicon Valley pour modèle et de donner sens à ce qui advient dans cette petite communauté (« la multitude » dans le lexique de l'État plateforme, ou, plus spécifiquement, des participants membres de l'administration qui mirent en œuvre un dédoublement critique). L'événement fut aussi une manière de neutraliser les conflits qui pouvaient exister autour du travail nécessaire aux informaticiens de l'administration fiscale pour publiciser le M ou encore de l'archaïsation dont ils firent les frais pour ne pas programmer l'impôt en Python ni déposer le code sur GitHub, perçus en 2016 comme étant deux standards incontournables[20].

20. Ces conflits ont été explicités a posteriori, dans un article du *Canard enchaîné* [10], ainsi que dans un tract de la CGT Finances publiques [17].

Enfin, examinons la citoyenneté comme *vérification* de l'activité de l'administration. En 2019, Denis Merigoux, alors doctorant en informatique dans un laboratoire de l'Inria spécialisé dans les méthodes formelles (la vérification mathématique des propriétés du langage, des étapes de compilation et de l'exécution d'un programme) décida d'appliquer les savoir-faire de son domaine de recherche au code M, désormais disponible en ligne. Intrigué par les enjeux logiques de la traduction du droit (fonctionnant par l'exposition successive des cas généraux, puis la liste des exceptions) dans un programme informatique dont l'exécution relève de l'arithmétique (le calcul de l'impôt), il commença une véritable enquête dans les plis de l'informatique fiscale. Il mit d'abord à jour des fichiers manquants dans le code source publié et des choix de simplification dans le compilateur vers Python au sujet desquels il s'adressa à Étalab, puis à la DGFIP. De proche en proche, il identifia des énigmes qui se détachaient de la trame du calcul et captèrent son attention : les choix d'arrondis, le calcul en nombres flottants, la double liquidation, etc. Les indices qu'il collectait au fil des échanges avec les administrations et de l'exercice du droit d'accès aux documents administratifs supposaient de recourir à des méthodes mises en œuvre dans le travail scientifique et de solliciter des collègues : son directeur de thèse, Jonathan Protzenko, un collègue doctorant à la Sorbonne, Raphaël Monnat avec qui il développa un compilateur, nommé MLang [16], mais aussi la juriste Liane Huttner, avec qui il conçut un langage dédié à la programmation du droit — qui a déjà fait l'objet d'une publication dans ce *Bulletin* [15] et lui valut le prix de thèse Gilles Kahn[21] décerné par la SIF en 2022. Pour le dire dans les termes du sociologue Luc Boltanski[22], le travail de ces chercheurs consistait à examiner ce qu'il en est de l'État, c'est-à-dire à faire travailler les inquiétudes, par exemple, auscultant, grâce à l'analyse critique du code, l'écart entre la lettre du droit ou l'intention du législateur et l'effectuation informatique du droit. Cela revenait donc, toujours dans les termes de Luc Boltanski, à ouvrir la possibilité que l'État se trouve mis en crise dans sa capacité à dire *la réalité de la réalité*.

Témoignage, valorisation, vérification désignent trois accomplissements pratiques du droit d'accès au document administratif qu'est un code source. Plutôt que de les considérer comme étant séparés, il faut retenir que ces exercices se sont adossé les uns aux autres et se complétaient. C'est parce qu'un étudiant a formulé, en 2014, une demande d'accès, qu'Étalab, qui n'aurait

21. https://archive.socinfo.fr/2022/12/recherche-prix-de-these-gilles-kahn-laureats-2022/.

22. Dans *Énigmes et complots* [4], Luc Boltanski retrace la manière dont l'État-nation détenteur du pouvoir d'organiser et d'unifier la réalité doit faire face, à partir de la fin du XIXe siècle et au début du XXe siècle, à l'émergence de nouvelles façons de mettre en doute la réalité. Énigmes, complots et enquêtes se multiplient à la fois dans la littérature populaire (naissance du roman policier, puis d'espionnage) et le journalisme, dans la psychiatrie clinique qui isole une nouvelle pathologie, la paranoïa, et du fait de l'apparition d'une nouvelle science de la société, la sociologie, qui peut se trouver en concurrence dans la manière d'établir les régularités qui encartent un événement. Autrement dit, la prétention de l'État à décrire la réalité est soumise à l'épreuve de capacités critiques, qui se développent dans diverses strates de la société et font évoluer les attentes en matière de gouvernement.

diplomatiquement pas pu agir contre une autre administration [19, p. 858], a organisé, en 2016, un hackathon de valorisation du code M de la DGFIP. De même, c'est parce que le code avait été publié à partir de 2016 et qu'il avait fait l'objet une communication sonore à l'occasion du hackathon, que Denis Merigoux en a eu connaissance et a mené, à partir de 2019, les vérifications nécessaires à la lecture du M, qui avaient manqué aux autres acteurs avant lui.

Étendre la citoyenneté à l'informatique, multiplier les enquêtes

Ces trois modalités d'exercice du droit d'accès correspondent à celles observées dans un cas spécifique, mais je formule l'hypothèse qu'elles pourraient constituer des points de comparaison utiles pour d'autres situations. Revenons aux logiciels participant des cadres juridico-informatiques de l'expérience de laboratoire, comme les logiciels de gestion de la paie, des congés ou des missions.

En citant l'article L300-2 du code des relations entre le public et l'administration,[23] notre chercheuse peut, depuis son bureau, faire une demande de droit d'accès à la direction des systèmes d'information de son institution de rattachement ou, si celle-ci s'avérait infructueuse, s'adresser à la CADA, puis au tribunal administratif. Le seul fait que la direction des systèmes d'information prenne conscience que ses logiciels sont des documents publics pouvant être communiqués à leurs utilisateurs (il y a de fortes chances qu'elle l'ignore !) pourrait déclencher un processus vertueux de contrôle que les récentes évolutions du droit aient bien été prises en compte, de hiérarchisation des opérations de maintenance, le cas échéant, de reprise en main de la délégation au prestataire, ou, a minima, engendrer un effort de documentation en vue de la transmission. Autrement dit, avant même de recevoir le code de ces programmes, *le seul fait d'en demander l'accès peut déclencher des opérations réflexives faisant progresser la conscience du droit*, de part et d'autre de l'interface. Voilà donc une vertu (à peu de frais pour la chercheuse) de l'exercice de témoignage de nature à piquer l'intérêt pour ces choses ennuyeuses que sont les codes des programmes que nous utilisons tous couramment.

Imaginons maintenant qu'à la suite de cette transmission, la direction des systèmes d'information de l'institution qui emploie la chercheuse souhaite valoriser l'effort de publication du code source des logiciels qu'elle utilise pour réaliser des traitements administratifs. Elle pourrait ainsi *déposer sur un répertoire public*, comme code.gouv[24] ou Software Heritage[25], les différentes versions des codes sources et permettre à d'autres de *les réutiliser et de contribuer à leur*

23. https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000033218936/2016-10-09.
24. https://code.gouv.fr/fr/.
25. https://www.softwareheritage.org/?lang=fr.

amélioration. Cette démarche pourrait également permettre à l'institution de s'inscrire non seulement dans le sillage du programme d'État plateforme et des démarches de transparence administrative, mais aussi d'appliquer à elle-même les attentes de science ouverte et d'archivage qui s'imposent aux publications des chercheurs.

Enfin, si la chercheuse décidait, par exemple, de travailler avec les syndicats à la vérification des codes sources, elle pourrait s'assurer que les primes aient bien été calculées et que l'ensemble des droits soient respectés dans la configuration du programme. Cette démarche de vérification formelle l'amènerait à remonter, comme Denis Merigoux l'a inauguré, une partie du processus d'altération informatique du droit, puisqu'elle devrait repartir de l'énoncé de droit pour vérifier si les règles se retrouvent effectivement dans le code. Il y a fort à parier que, dans le long processus de traduction des règles de droit vers des programmes supposant de décliner les textes dans le contexte de l'institution, de sélectionner les parties relatives au calcul, de spécifier pour le programme et de programmer, avant d'exécuter [1], outre de potentielles erreurs, elle découvrirait l'existence de microchoix de programmation, dont la direction des systèmes d'information n'a pas toujours conscience. Ainsi, dans un récent travail sur le calcul des aides au logement par la Caisse nationale d'allocations familiales, mené avec Denis Merigoux et une collègue juriste, Lilya Slimani, nous avons identifié des imprécisions par rapport au texte de droit, une tentation de simplifier les catégories (certes, souvent retorses !), et des problèmes d'invisibilité des cas limites [14]. Cette vérification de la traduction du droit peut *aider les administrations à limiter l'écart entre la réalité et l'ordre déclaré* — soit parce qu'elles auraient acquis des prises pour se réformer elles-mêmes, soit parce que la chercheuse et les syndicats disposeraient d'appuis juridiques suffisamment précis pour opposer, dans une démarche contentieuse, le droit à l'État.

Dans cette adresse aux membres de la Société d'informatique de France, mon propos était de formuler un plaidoyer en faveur d'un exercice de la citoyenneté étendu aux cadres juridico-informatiques de l'expérience sociale, et d'attirer l'attention des communautés de recherche qui disposent de savoir-faire en matière de lecture et de rétro-ingénierie de programmes, sur le fait que cette expertise peut être mise à profit de toute la société politique. Afin de souligner une dernière fois les enjeux d'une telle cause, j'insiste sur deux éléments déterminants.

Premièrement, même si le droit d'accès est garanti depuis près de 50 ans par le droit national, *son application est très fréquemment entravée*. L'obstruction de l'administration n'est pas toujours intentionnelle. Elle mêle une pluralité de motifs : la demande peut ne pas être pas arrivée aux bons responsables, les agents et leur hiérarchie peuvent méconnaître le droit, ne pas savoir comment l'appliquer au cas d'espèce et préférer s'abstenir. Surtout, le coût pour rechercher, sélectionner ou caviarder les documents avant de les transmettre peut

paraître élevé aux administrations qui n'ont pas l'habitude de traiter ce type de demande. C'est très largement le cas pour les documents administratifs communicables que sont les codes sources. Ainsi, si l'on parcourt le site Ma Dada, conçu par l'Open Knowledge Fondation pour faciliter les demandes d'accès, on y lit, à l'été 2024, une demande d'accès au code et à la documentation de l'application Rdv étrangers, utilisée dans les préfectures pour les demandes de titres et pour laquelle le ministère de l'Intérieur a déjà un retard supérieur au délai légal de 30 jours[26] ; une autre, refusée, pour accéder au code de l'application de signalement des « anomalies » sur l'espace public, Proxim'Cité, de la métropole de Clermont-Ferrand[27], ainsi que de nombreuses demandes ayant mené à une saisine de la CADA ou à un recours gracieux[28] (y compris celle concernant la plateforme bien connue des universitaires, Galaxie)[29]. Il est donc impératif d'exercer régulièrement ce droit pour que les agents publics améliorent leur capacité à rendre des comptes sur leur administration et réalisent leur mission en gardant aux marges de leur conscience le fait qu'ils travaillent dans et pour le compte d'une société démocratique vigilante.

Deuxièmement, peu d'autres groupes professionnels que les chercheurs en informatique me semblent plus indiqués pour l'exercice de cette citoyenneté informatique, et ce pour plusieurs raisons. Vous disposez des *compétences techniques* pour mettre en place des démarches de vérification de programmes (souvenons-nous qu'Étalab, qui emploie de nombreux informaticiens, n'a pas su exploiter un code écrit dans un langage dédié). Le travail à l'université et dans les instituts de recherche facilite la rencontre avec des collègues d'autres disciplines susceptibles d'accompagner une telle démarche, à commencer par les collègues juristes dans les facultés de droit. En tant qu'agents exerçant également des missions de service public, vous partagez une *disposition pour la vie publique* qui n'est pas forcément valorisée chez les informaticiens d'un laboratoire industriel de recherche et développement ou travaillant dans une entreprise de service numérique. En outre, par vos enseignements, vous pouvez participer à façonner la conscience citoyenne des générations suivantes, notamment en prêtant attention aux exemples utilisés en cours et en étant soucieux d'inscrire dans les formations des enseignements consacrés à la programmation, à la compilation et à la vérification des systèmes. Pour toutes ces raisons, les chercheurs en informatique des universités, de l'Inria et du CNRS pourraient multiplier les sites d'investigation et devenir les fers de lance de cette citoyenneté informatique qui apparaît de plus en plus essentielle.

26. https://madada.fr/demande/code_source_de_lapplication_rdv#outgoing-49223.
27. https://madada.fr/demande/code_source_de_loutil_proximcite#incoming-9724.
28. https://madada.fr/demandes/toutes?utf8=%E2%9C%93&query=%22code+source%22&request_date_after=&request_date_before=&commit=Chercher.
29. https://madada.fr/demande/code_source_du_logiciel_galaxie#outgoing-48985.

Remerciements

Ce texte poursuit une réflexion présentée dans le séminaire Codes sources du 4 avril 2024[30] à l'invitation d'Emmanuel Chailloux (LIP6, IRILL, Sorbonne Université), de Raphaël Fournier-S'niehotta (CNAM) et de Baptiste Mélès (CNRS, Archives Henri-Poincaré, ambassadeur Software Heritage). Je les remercie, ainsi que Louis Gesbert, Denis Merigoux, Florence Renucci, Didier Torny, et tous les participants de leurs commentaires et réactions.

Références

[1] Marie Alauzen. à paraître en 2024. L'altération informatique du droit. Une sociologie du passage du droit aux droits. *Droit & Société* 2, 117. Retrieved from https://hal.science/hal-04650848.

[2] Marie Alauzen and Fabian Muniesa. 2022. Attention aux mots d'ordre ! Éléments pour une enquête sur la signification de l'État. *Tracés : Revue de Sciences Humaines* 43, 43: 35–54. https://doi.org/10.4000/traces.14146.

[3] Philippe Bezes. 2009. *Réinventer l'État. Les réformes de l'administration française (1962-2008)*. Presses Universitaires de France, Paris.

[4] Luc Boltanski. 2012. *Énigmes et complots : Une enquête à propos d'enquêtes*. Gallimard.

[5] Geoffrey Bowker and Susan Leigh Star. 2000. *Sorting things out: Classification and its consequences*. MIT Press, Cambridge, MA.

[6] Erving Goffman. 1991. *Les cadres de l'expérience*. éd. de Minuit (Coll. « Sens commun »), Paris.

[7] Frédéric Graber. 2022. *Inutilité publique. Histoire d'une culture politique française*. Amsterdam, Paris.

[8] Frédéric Graber. 2023. *L'affichage administratif au XIXe siècle. Formuler le consentement*. Presses de la Sorbonne, Paris.

[9] Léo Griffaton-Sonnet. 2017. Quelle portée pour l'article 15 de la déclaration des droits de l'homme et du citoyen ? *Revue française de droit constitutionnel* 112, 4: 899–918. https://doi.org/10.3917/rfdc.112.0899.

[10] Alain Guédé. 2016. Pression fiscale à "c dans leurre". *Le Canard enchaîné*: 4.

[11] Lilly Irani. 2015. Hackathons and the making of entrepreneurial citizenship. *Science, Technology, & Human Values* 40, 5: 799–824. https://doi.org/10.1177/0162243915578486.

[12] Martha Lampland and Susan Leigh Star. 2009. *Standards and their stories: How quantifying, classifying and formalizing practices shape everyday life*. Cornell University Press, Ithaca, NY.

[13] Dominique Linhardt. 2009. L'État et ses épreuves. Élements d'une sociologie des agencements étatiques. *Clio@Thémis : Revue électronique d'histoire du droit*, 1: 36pp. Retrieved from https://doi.org/10.35562/cliothemis.1978.

[14] Denis Merigoux, Marie Alauzen, and Wouroud Slimani. 2024. Rules, computation and politics: Scrutinizing unnoticed programming choices in french housing benefits. *Journal of Cross-Disciplinary Research in Computational Law* 2. Retrieved from https://journalcrcl.org/crcl/article/view/32.

30. https://codesource.hypotheses.org/1256.

[15] Denis Merigoux and Liane Huttner. 2022. Catala : Formaliser la loi grâce à un langage de programmation dédié. *1024* 20: 77–84. https://doi.org/10.48556/SIF.1024.20.77.

[16] Denis Merigoux, Raphaël Monat, and Jonathan Protzenko. 2021. A Modern Compiler for the French Tax Code. In *CC '21: 30th ACM SIGPLAN International Conference on Compiler Construction* (CC 2021: Proceedings of the 30th ACM SIGPLAN international conference on compiler construction), 71–82. https://doi.org/10.1145/3446804.3446850.

[17] CGT Finances publiques. 2016. L'informatique à la DGFIP ? Ce n'est pas qu'une affaire d'informaticiens ! *Tract*.

[18] Pierre Rosenvallon. 2004. *Le modèle politique français : La société civile contre le jacobinisme de 1789 à nos jours*. Seuil, Paris.

[19] Sébastien Shutz. 2019. Un logiciel libre pour lutter contre l'opacité du système socio-fiscal. Sociologie d'une mobilisation hétérogène aux marges de l'État. *Revue française de science politique* 69, 5-6: 845–868.

[20] Susan Leigh Star. 1998. Grounded classifications: Grounded theory and faceted classifications. *Library Trends* 47: 218–32.

[21] Susan Leigh Star. 1999. The ethnography of infrastructure. *American Behavioral Scientist* 43: 377–91.

[22] Susan Leigh Star and Anselm L. Strauss. 1999. Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer Supported Cooperative Work: The Journal of Collaborative Computing* 8: 9–30.

Effets environnementaux de la 5G (partie 2)

Applications envisagées et acteurs impliqués

Philippe Ciblat[1], Jacques Combaz[2], Marceau Coupechoux[3], Kevin Marquet[4], Anne-Cécile Orgerie[5]

Cette série d'articles a pour objectif de rassembler les connaissances actuelles liées aux effets environnementaux de la 5G. Elle s'articule autour de questions liées à ces effets, et est organisée en trois parties : la première (publiée dans le numéro 23 de ce bulletin, [31]) donnait les bases nécessaires à la compréhension de la technologie elle-même ; la deuxième, présentée ici, détaille les services rendus possibles par la 5G et les acteurs impliqués ; la dernière partie, enfin, se penchera sur les effets environnementaux du déploiement de la 5G (elle sera publiée dans le prochain numéro de ce bulletin).

Dans cette partie, nous décrivons les applications envisagées de la 5G et les stratégies des différents acteurs vis-à-vis de ces applications (consommateurs, États, opérateurs, fournisseurs de services, équipementiers, collectivités locales). Notre démarche est de décrire les applications permises par la 5G ainsi que les relations entre les différents acteurs de la 5G. Notre objectif est de

1. LTCI, Telecom Paris, IP Paris, EcoInfo.
2. Université de Grenoble Alpes, CNRS, Grenoble INP, VERIMAG, EcoInfo.
3. LTCI, Telecom Paris, IP Paris, LinCS, EcoInfo.
4. Université de Lyon, INSA Lyon, Inria, CITI, EcoInfo.
5. Université de Rennes, Inria, CNRS, IRISA, EcoInfo.

DOI:10.48556/SIF.1024.24.99

nous servir des descriptions de cet objet socio-technologique pour ensuite évaluer sa pertinence écologique (qui sera développée dans la partie 3). Notre posture n'est pas de critiquer un acteur de la 5G en particulier, mais plutôt de pointer les intérêts divergents entre les différents acteurs.

À quelles applications va servir la 5G ?

Les spécifications de la 5G permettent d'envisager le déploiement de nouvelles applications. Elles décrivent trois grandes familles d'usages et d'applications que nous passons en revue dans cette section. Il faut noter que chaque catégorie d'usage correspond à des modes de fonctionnement du réseau pas nécessairement compatibles entre eux. C'est-à-dire que les niveaux de performances ne peuvent pas être maximaux simultanément dans tous les domaines (en débit, en latence, etc.) et ces modes de fonctionnement peuvent correspondre à des configurations différentes du réseau, notamment en termes de déploiement d'infrastructures (nombre d'antennes, portée des antennes, etc.).

a. **L'*Enhanced Mobile Broadband (eMBB)*** exploite principalement la *bande passante élevée*, et se destine, entre autres :
   – au *streaming* vidéo haute définition (4K) voire volumétrique (images immersives en 3D) ;
   – à la réalité virtuelle et à la réalité augmentée, qui profitent également d'une latence réduite ;
   – aux jeux en réseaux (ou *cloud gaming*), qui nécessitent également une latence réduite ;
   – à la télémédecine, qui inclurait par exemple les téléconsultations et les communications entre l'ambulance et l'hôpital ;
   – les communications depuis des transports à haute vitesse, car la 5G doit offrir une meilleure gestion du changement de station de base nécessaire lors de la mobilité des équipements.

   Afin de fournir des débits plus importants, l'eMBB s'appuie notamment sur des largeurs de bande plus grandes qu'en 4G (notamment dans les bandes millimétriques) et sur la technologie d'antennes multiples MIMO massifs (voir section « Comment (technologiquement) va-t-on atteindre ces performances ? » de la première partie de ce rapport [31]).
b. **L'*Ultra-Reliable and Low Latency Communications (uRLLC)*** concerne principalement les communications à *faible latence* et *ultra fiables*, dont des applications potentielles sont :
   – les véhicules autonomes, auxquels la 5G doit offrir des communications fiables avec de très faibles latences afin de déporter dans le cloud des calculs coûteux de prédiction de trajectoire [28], ou d'assurer des communications rapides avec l'infrastructure routière pour prévenir des accidents [74], ou encore des communications rapides entre véhicules afin

de coordonner de manière sécurisée des changements de voie par exemple [21] ;

– l'industrie 4.0, une notion recouvrant différentes significations suivant les contextes [40]. Elle peut désigner une chaîne de production dans laquelle toutes les entités sont équipées de capacités communicantes pour faire remonter des informations ou être pilotées à distance à des fins d'optimisation. Ces communications peuvent s'effectuer sur un réseau 5G public ou privé [77], [44] ;
– la téléchirurgie (chirurgie à distance), ou plus largement le domaine de la *e-santé* [84]. Il s'agit par exemple de permettre à un chirurgien spécialisé dans un type d'opération de commander les machines qui opèrent un patient potentiellement situé ailleurs dans le monde. Une opération chirurgicale du cerveau d'un patient, distant de plus de 3000 kilomètres du chirurgien, a été réalisée en 2020 [98]. Pour cette opération, les équipements du chirurgien à distance étaient connectés à un réseau 5G ainsi que les équipements robotisés opérant le patient. Il est à noter qu'entre les deux stations de base 5G reliant respectivement le patient et le chirurgien, un réseau filaire est utilisé (réseau du cœur de l'Internet), réseau qui n'est pas spécifique au déploiement de la 5G. La téléchirurgie utilisant la 5G à courte distance pourrait permettre de supprimer des câbles dans les hôpitaux. La 5G devrait permettre également des diagnostics à distance (téléconsultation) dans des zones rurales ou éloignées, manquant de médecins spécialisés [88]. Les réseaux satellites sont également censés assurer ce service [52]. Cependant, pour ce type de consultation à distance, des débits relativement faibles (de 318 à 512 kbps) sont suffisants d'après des études en conditions réelles [88]. Dans ces cas-là, le souci des réseaux actuels vient plutôt des déconnexions intempestives. Ce problème pourrait être corrigé en optimisant le déploiement des antennes 3G et 4G ou en donnant la priorité au trafic des applications médicales à la manière de ce qui est envisagé avec le *network slicing* pour la 5G. Enfin, dans le domaine de la santé, le déploiement de la 5G permettrait d'éliminer les câbles de communication dans le pilotage à distance des robots ;
– les applications militaires [22], telles que la reconnaissance du terrain ou le guidage de véhicules militaires grâce aux faibles latences et la grande fiabilité.

Les communications uRLLC s'appuient sur une redondance très importante de l'information transmise afin d'en assurer la fiabilité. Les protocoles de la 5G permettent d'envoyer un paquet de données quasiment dès son arrivée en mémoire, de transmettre des paquets toutes les 250 μs contre 1 ms en 4G, d'obtenir des acquittements de la part du récepteur beaucoup plus rapidement qu'en 4G. Les données critiques peuvent également préempter les ressources prises pour l'eMBB. Ces innovations techniques permettent d'offrir des

latences beaucoup plus faibles qu'avec les générations précédentes de protocoles de réseaux mobiles.

c. **Le *Massive Machine Type Communications (mMTC)*** exploite principalement le *grand nombre de connexions simultanées* et la *plus grande efficacité énergétique* permis par la 5G. Les applications principales du mMTC proviennent de l'*Internet des objets (IoT — Internet of Things)* avec la connexion d'un grand nombre d'objets (ce nombre varie suivant les applications) pour permettre la remontée de grandes quantités d'information, l'automatisation accrue de multiples processus, ou encore la surveillance des biens ou des personnes [61], [83]. Par exemple, dans le contexte de l'agriculture connectée, le contrôle à grande échelle des bâtiments et des installations permet d'assurer des opérations de logistique en temps réel ; le pilotage de flottes de drones agricoles permet de réaliser de l'agriculture de précision, etc. [89]. Les protocoles 4G ne sont pas adaptés à un trafic de données sporadique, pas nécessairement de fort débit, mais généré par un grand nombre de terminaux en même temps. Chaque connexion requiert par exemple des échanges de messages de signalisation qui peuvent congestionner le réseau si des centaines de terminaux tentent d'accéder au réseau simultanément.

Certains opérateurs motivent le passage à la 5G par le fait que cette technologie permettrait l'avènement de la ville intelligente, ou *smart city*. Ce concept décrit un territoire sur lequel «*des infrastructures numériques permettent aux entreprises de collaborer et d'innover, de fournir de meilleurs services aux habitants*» [59], ou encore sur lequel «*des problèmes sociaux ou environnementaux sont réglés grâce à des solutions numériques*» [41], sans qu'il y ait de définition consensuelle de ce concept. De nombreuses applications de la ville intelligente reposent sur la numérisation massive des infrastructures urbaines (bâtiments, transports, personnes, maisons, etc.). La 5G peut faciliter ces connexions massives d'objets hétérogènes [87], notamment grâce au mode mMTC.

Le réseau 5G n'est pas toujours indispensable pour réaliser certaines des applications citées précédemment. Par exemple, des véhicules autonomes sont actuellement en test et n'utilisent pas de réseau mobile. Le trafic de certaines applications requérant un fort débit peut être servi par le WiFi et ses évolutions. Un des intérêts du déploiement de la 5G pourrait être de réaliser toutes ces applications avec une seule infrastructure, plutôt que de multiplier les solutions applicatives dédiées (qu'elles soient basées sur des technologies existantes ou à développer).

Par ailleurs, les verrous technologiques de ces applications ne concernent pas que les performances ou la nature des communications sans fil. Ainsi, certains des paris technologiques cités précédemment supposent des progrès importants en matière de puissance de calcul ou d'efficacité énergétique, en matière de production et de stockage de l'énergie, etc. Il reste donc une large part

d'incertitude sur les nouvelles applications qui seront accessibles grâce au déploiement de la 5G.

Notons que les premiers déploiements de réseaux 5G en France ne visent que la catégorie *eMBB*. Les catégories *uRLLC* et *mMTC* sont ciblées par des vagues ultérieures de déploiements et d'améliorations du réseau mobile. En particulier, l'uRLLC nécessite des fonctionnalités dans le réseau d'accès et dans le réseau cœur qui ne sont pas encore déployées en France. Le réseau d'accès doit par exemple s'appuyer sur une trame TDD (utilisée pour le duplexage temporel [53]) adaptée aux faibles latences, or ce n'est pas le choix qui a été fait dans un premier temps. Un réseau cœur 5G est également nécessaire, alors que les premiers déploiements s'appuient sur le réseau cœur de la 4G (5G Non-Stand-Alone, voir section « Comment (technologiquement) va-t-on atteindre ces performances ? » de la première partie de ce rapport [31]). Enfin l'uRLLC nécessite énormément de ressources radio pour assurer fiabilité et faible latence. Les simulations du 3GPP montrent par exemple que dans un scénario d'automatisation d'usine, une cellule ne peut supporter plus de 10 terminaux exigeant une fiabilité de 99,9999 % et une latence d'une milliseconde, dans 40 MHz de bande passante pour un débit total ne dépassant pas 1,3 Mbps [6].

Le développement concomitant d'applications des différentes catégories eMBB, uRLLC et mMTC sur une même infrastructure physique nécessite le déploiement du *network slicing* (voir section « Comment (technologiquement) va-t-on atteindre ces performances ? » de la première partie de ce rapport [31]). Or cette technologie n'est pas encore bien maîtrisée au moment des premiers déploiements de la 5G et pourrait être freinée par le principe de neutralité du Net qui interdit aux opérateurs de traiter les types de trafic de manière différenciée. Ces enjeux de neutralité du Net constituent un des sujets récurrents de controverse autour de la 5G [30], à la croisée entre enjeux géopolitiques et économiques [45]. L'Union européenne a en effet adopté en 2015 le règlement nº 2015/2110 garantissant la neutralité du Net au sein de l'Union [1], tandis que les États-Unis ont mis fin à ce principe sur leur territoire en 2017 [45] avant de le restaurer en 2024 [47]. D'autres pays proposent déjà des offres 5G enfreignant le principe de neutralité du Net. Ericsson a conduit une étude en 2023 sur 105 fournisseurs d'accès 5G dans 28 pays qui indique que 13 % de ces fournisseurs ont commencé à mettre en œuvre des offres impliquant différentes qualités de service [86]. Ils citent notamment les exemples suivants : le forfait *5G Stock Pro* de 3HK [7] à Hong Kong qui alloue davantage de ressources et donne la priorité sur l'utilisation du réseau pour des applications boursières de gestion d'actions ; China Unicom en Chine qui propose un forfait *Super Live Streaming* qui donne la priorité à la liaison montante pour du streaming vidéo live ; et le service *ProPing* d'Ooredoo [67] au Koweït qui garantit une faible latence pour les jeux mobiles sur la 5G [86]. Ces offres ciblent des applications spécifiques, déjà permises par la 4G, mais dont la 5G devrait améliorer les performances et dans les trois cas cités, cette amélioration de

performances implique un traitement différencié du trafic de ces applications : plus de ressources réseau, trafic prioritaire ou faible latence garantie. Ces traitements font l'objet d'offres commerciales spécifiques de la part des fournisseurs 5G concernés : l'amélioration de la qualité de service repose sur des tarifs différents.

Les applications susmentionnées sont celles envisagées par les opérateurs, les fournisseurs de service ou les équipementiers. Cependant, « *l'histoire des techniques montre qu'il y a toujours un écart entre les prévisions des concepteurs et les pratiques de celles et ceux qui s'approprient les technologies, souvent de manière très inattendue, voire transgressive* » [30]. Il est ainsi très difficile de prévoir les applications qui seront réellement déployées, en particulier concernant les *nouvelles* applications qui seraient rendues possibles par la 5G.

Qui va bénéficier du déploiement de la 5G ?

Différents acteurs jouent un rôle important dans le déploiement des infrastructures de la 5G :

– les consommateurs et consommatrices qui utilisent les infrastructures ;
– l'État qui conditionne, notamment financièrement, le déploiement ;
– les opérateurs qui déploient et gèrent les infrastructures ;
– les fournisseurs de services dont les services en question vont générer les données circulant sur les infrastructures ;
– les équipementiers qui fabriquent et commercialisent les équipements nécessaires au déploiement de la 5G ;
– les collectivités locales où sont implantées les infrastructures.

Le dimensionnement des infrastructures doit être adapté aux usages, et inversement. L'ampleur du déploiement et de l'usage d'infrastructures relativement lourdes comme celle du réseau 5G dépend donc en grande partie des motivations de ces différents acteurs. Nous donnons dans la suite quelques clefs de compréhension des motivations et du comportement de ces acteurs par rapport au déploiement de la 5G.

Consommateurs

Les consommateurs et consommatrices devraient bénéficier des nouvelles capacités du réseau mobile et des nouveaux services proposés, mais de façon différenciée en fonction des territoires et des logiques de déploiement qui seront à l'œuvre (voir la section « Faut-il différencier les besoins en 5G selon les territoires ? »). En effet, les réseaux 5G déployés dans le monde diffèrent considérablement en performances et taux de couverture en fonction des options choisies par les États et opérateurs (bandes de fréquence, densité des équipements, déploiement géographique, etc.).

Remarquons qu'en France le déploiement de la 5G a fait l'objet en 2020 d'une demande de moratoire portée par la Convention citoyenne pour le climat et par certains élus [79], [48]. On peut s'interroger sur l'hypothèse que ce nouveau réseau corresponde à une attente *a priori* des consommateurs. On peut également s'interroger sur les nouvelles capacités offertes par le réseau et leur potentiel (associé à la publicité et au marketing) à créer de nouveaux besoins chez les consommateurs et consommatrices. C'est une des préoccupations fortes des citoyens et citoyennes de la ville de Rennes qui ont participé à la mission d'étude sur la 5G pour éclairer le débat public [62].

État

Fin 2020, l'État français a mis aux enchères des blocs de fréquences dédiées à la 5G à destination des opérateurs français [3]. Avant ces enchères, les opérateurs Orange, SFR, Bouygues Telecom et Free ont chacun obtenu un bloc de 50 mégahertz (MHz) dans la bande de fréquences aux alentours de 3,5 GHz pour un prix fixé de 350 millions d'euros par bloc. Lors des enchères, les opérateurs se sont partagé 11 blocs supplémentaires de 10 MHz. Cette opération a rapporté près de 2,8 milliards d'euros à l'État français [58]. À titre de comparaison, le Royaume-Uni a également vendu aux enchères des blocs de fréquences dédiés à la 5G en 2021 aux opérateurs pour un montant total de 1,36 milliards de livres sterling [66].

L'ARCEP s'est appuyé sur trois méthodes d'évaluation [33] pour fixer le prix de réserve (prix initial) des blocs de 10 MHz mis aux enchères :

– le parangonnage : évaluation fondée sur le prix des fréquences 3,5 GHz attribuées pour des services mobiles dans d'autres pays, en tenant compte du pouvoir d'achat des habitants ;
– la valeur actuelle nette : évaluation des flux de trésorerie générés par l'exploitation des fréquences concernées pour un opérateur générique ;
– la méthode économétrique : analyse économétrique sur cent licences attribuées aux opérateurs de téléphonie mobile depuis 2003 afin de rechercher les facteurs explicatifs du prix des licences.

La deuxième méthode requiert l'évaluation de nombreux paramètres dont le volume de trafic prévisionnel sur la durée de l'allocation des fréquences, à savoir 15 ans. Elle s'appuie ainsi sur un scénario de croissance en suivant une fonction logistique calibrée sur les consommations actuelles de données et qui prévoit une consommation mensuelle moyenne par carte SIM en 2034 de 55 à 65 Go par mois [33].

Au niveau européen, une des principales préoccupations des États concerne la sécurité des infrastructures et en particulier les problématiques de cybersécurité liées aux fabricants d'équipements 5G [34], [76]. Le parlement européen a

rappelé en 2019 les enjeux de souveraineté européenne dans ce domaine et a voté une résolution pour mettre en garde les États membres sur les menaces de sécurité liées aux équipements vendus par les entreprises chinoises Huawei et ZTE [75]. L'Union Européenne, comme dans d'autres secteurs industriels, soutient financièrement la recherche et le développement (R&D) et l'innovation des entreprises dans le domaine de la 5G, notamment Nokia et Ericsson qui sont des fabricants d'équipements respectivement finlandais et suédois [26]. De manière similaire, les États-Unis soutiennent financièrement les concurrents de Huawei, «*cible privilégiée de la guerre commerciale entre Washington et Pékin*» [45].

Opérateurs

Pour les opérateurs (ou FAI, fournisseurs d'accès à internet), en plus du prix susmentionné, l'acquisition des licences s'accompagne d'obligations de déploiement de stations de base 5G pour chaque opérateur, fixées par l'ARCEP dans la bande 3,4–3,8 GHz : 3 000 stations en 2022, 8 000 en 2024 et 10 500 en 2025 [11]. Le déploiement de la 5G nécessite donc de lourds investissements de la part des opérateurs. D'après la Fédération française de télécom, les investissements dans la 5G s'élèvent à 1,5 à 2,2 milliards d'euros par an [65]. Il est donc crucial pour les opérateurs que ces investissements soient rentabilisés.

Afin d'apprécier l'ampleur de l'hypothèse de croissance faite par l'ARCEP pour fixer le prix de réserve des blocs de fréquences et l'impact qu'elle a sur les opérateurs, observons la figure 1. Elle montre la consommation mensuelle moyenne par carte SIM en France de 2009 à 2023 hors cartes MtoM (cartes *Machine to Machine* qui sont présentes par exemple dans les systèmes d'alarmes, les ascenseurs, les caméras de sécurité, les panneaux d'affichage, etc.). Les données sont issues des rapports annuels de l'observatoire des marchés des communications électroniques en France de l'ARCEP [19], ces données sont disponibles dans l'annexe A. Cette consommation par carte SIM, tout comme la consommation globale de données, suit une croissance linéaire depuis quelques années, après plusieurs années de croissance exponentielle (la section «La 5G va-t-elle induire une augmentation du trafic?» détaille ce point). La courbe bleue montre une régression linéaire sur les données de 2016 à 2023 avec en pointillés, la projection de la courbe jusqu'à 2034. On observe que la fourchette retenue dans l'évaluation du prix de réserve des blocs de fréquence (55 à 65 Go par mois comme indiqué précédemment) est très nettement supérieure à cette projection (illustrée en figure 1); les analystes misent donc sur une croissance du volume de trafic par carte SIM nettement plus rapide que ce qu'elle est actuellement.

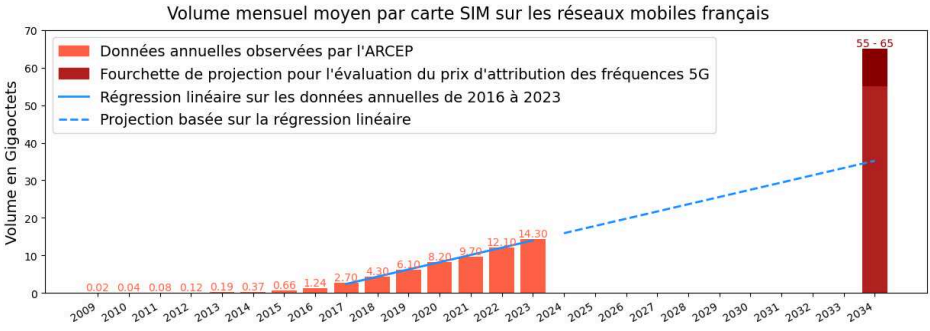


Fig. 1. Données consommées en moyenne par mois et par carte SIM sur les réseaux mobiles français hors cartes MtoM (données ARCEP) et projection pour 2034.

De là, il nous faut commenter l'existence de plusieurs effets contradictoires sur les motivations des opérateurs à déployer la 5G. D'un côté, l'acquisition aux enchères de nouvelles fréquences dédiées à la 5G représente un investissement lourd pour les opérateurs qui, de surcroît, s'accompagne d'obligations de déploiement coûteuses elles aussi. Ces investissements conséquents peuvent entraîner des risques conséquents sur la viabilité économique des opérateurs, voire dans certains cas un resserrement du nombre d'opérateurs par pays et donc des situations renforcées de monopole économique [60], [26]. D'un autre côté, certaines analyses soulignent ces dernières années que les opérateurs ont besoin de s'engager dans le déploiement de la 5G afin de compenser le *revenue gap* : le trafic en forte augmentation génère des frais liés à la consommation électrique notamment, alors que les revenus n'augmentent que faiblement [63], [81], [80], [64], [91]. L'augmentation importante de l'efficacité énergétique serait ainsi une condition nécessaire à la viabilité économique des opérateurs. En outre, comme pour chaque nouvelle génération de communications mobiles, le contexte fortement concurrentiel rend nécessaire l'engagement des opérateurs dans le déploiement de la dernière génération de réseau mobile [73]. Des revenus supplémentaires pourraient être obtenus via le développement de nouveaux services fournis par les opérateurs, par exemple dans les domaines évoqués dans la section précédente (véhicules autonomes, télémédecine, industrie 4.0, etc.). De tels nouveaux services pourraient accélérer l'augmentation du trafic de données telle qu'elle a été supposée dans le calcul du prix de réserve d'attribution des fréquences aux enchères (figure 1). Cependant, ces nouveaux cas d'utilisation sont encore en cours de développement et devront trouver des consommateurs pour générer des revenus pour les opérateurs [26].

Dans ce contexte, les opérateurs appliquent des stratégies économiques variées pour rentabiliser leurs investissements. Ces stratégies ont un impact

réel sur les déploiements de la 5G. Pour l'illustrer, prenons les cas de Free et d'Orange. En 2021, Free n'avait pas augmenté le tarif de ses abonnements avec l'arrivée de la 5G, si bien que Free proposait des tarifs inférieurs aux forfaits concurrents d'Orange tout en offrant des volumétries supérieures. En 2021, Free avait déployé beaucoup plus de sites 5G qu'Orange, mais il s'agissait surtout de sites à 700 et à 800 MHz [13]. À l'époque, Free était d'ailleurs le seul à avoir fait le choix des fréquences 700 et 800 MHz pour ses sites 5G, les trois autres principaux opérateurs français s'étant tournés vers des fréquences de 1,8 et 2,1 GHz, en plus de la fréquence 3,5 GHz commune à tous les opérateurs. En 2022, d'après l'ARCEP [12], les performances du réseau 5G Free étaient inférieures à celle du réseau 5G Orange en termes de débit (montant et descendant).

Fournisseurs de services

Si certains services sont rendus, et donc monétisables, par les opérateurs, l'augmentation du trafic est principalement liée à l'évolution des services en ligne proposés par d'autres acteurs (les fournisseurs de services), qui génèrent des données circulant sur le réseau (cf. section « À quelles applications va servir la 5G ? »). L'ARCEP estime que fin 2022, environ 54 % du trafic vers les clients (mobiles et fixes confondus) des principaux FAI (fournisseurs d'accès à Internet) en France (Bouygues, Free, Orange, SFR) provient des 5 acteurs suivants : Netflix, Google, Akamai, Meta et Amazon [17]. Selon l'ARCEP : *« Ceci indique une concentration de plus en plus nette du trafic autour de quelques acteurs »* [17]. Outre la concentration du trafic autour d'une poignée de fournisseurs de services, on peut noter que l'explosion du trafic lié à ces services est principalement imputable à du trafic vidéo, notamment à du *streaming* (à la demande et live). En effet, les 5 acteurs mentionnés fournissent massivement des contenus vidéos : Netflix, Google (avec YouTube), Akamai (avec notamment Disney+ en France), Meta (avec Facebook) et Amazon (avec Prime) [17]. Au niveau mondial, le constat est similaire : Sandvine estimait que, sur la première moitié de l'année 2022, le trafic vidéo comptait pour 65,93 % dans le trafic IP mondial (tous réseaux d'accès confondus) et pour 67,60 % dans le trafic mobile mondial [82].

La deuxième catégorie d'applications (après la vidéo) la plus contributrice en termes de volume de données sur les réseaux mobiles est la catégorie des réseaux sociaux (Facebook, Instagram, Snapchat, etc.) au niveau mondial selon Sandvine [82]. Ces deux types d'applications, streaming vidéo et réseaux sociaux, ont en commun de faire reposer leur viabilité financière en grande partie sur des mécanismes liés à l'économie de l'attention [29]. Ces mécanismes datent certes du XIX^{e} siècle, mais ils n'ont jamais été autant efficaces qu'à l'heure actuelle [96]. En effet, au-delà des revenus générés par les abonnements pour certains de ces services, les publicités et en particulier les publicités

ciblées (et donc la collecte de données personnelles des consommateurs et consommatrices) sont une source majeure de revenus pour les fournisseurs de services. La lecture automatique de vidéos, le défilement infini, les récompenses sociales ou non sociales, sont autant de mécanismes visant à capter ou retenir l'attention des utilisateurs et utilisatrices [29] et par ricochet, à générer plus de trafic.

La quantité de données générées par ces services en ligne, tel que le *streaming* vidéo, est donc très importante et requiert de la part des opérateurs un effort d'investissement accru dans le déploiement de leurs infrastructures, notamment mobiles. Si le projet de faire participer certains fournisseurs de service à cet effort a été évoqué ces dernières années en France [35], il n'est pour le moment pas d'actualité. Notons enfin que la frontière entre opérateurs (FAI) et fournisseurs de services devient de plus en plus floue [45]. Par exemple Google, qui dispose d'un important réseau privé (notamment des fibres sous-marines), a commencé il y a plusieurs années à proposer des forfaits *Google Fiber* d'accès à Internet via la fibre aux États-Unis [50] et commercialise également depuis plusieurs années des smartphones (*Google Nexus* et *Google Pixel*), dont les derniers modèles supportent la 5G [51].

Équipementiers

Comme indiqué précédemment, les équipementiers liés à la 5G font l'objet d'enjeux complexes, à la fois géopolitiques, économiques, de souveraineté et de cybersécurité, avec d'un côté les entreprises chinoises (Huawei et ZTE) et de l'autre les entreprises européennes (Ericsson et Nokia) [45], [34], [76], [26], [46]. Ces équipementiers fournissent les nouvelles stations de base 5G. Leurs profits potentiels sont considérables si le déploiement de la 5G s'avère conséquent. Ces gains potentiels concernent aussi les équipementiers du cœur de réseau qui fournissent des routeurs supportant la 5G tels que l'américain Cisco. Ces différents équipementiers publient très régulièrement des rapports de prospective prédisant des augmentations importantes : du trafic mobile à venir, du nombre d'applications supportées par la 5G, du nombre d'utilisateurs et utilisatrices ou encore du nombre d'équipements connectés aux réseaux mobiles (voir par exemple : Cisco [32], Ericsson [86], Huawei [42], Nokia [49]). De nombreux lobbies tentent d'influencer les États et les FAI pour un déploiement massif de la 5G, amenant ainsi certains économistes à citer la 5G comme un cas classique de *technology push* : une forte pression pour le déploiement d'une nouvelle technologie sans demande claire des utilisateurs et utilisatrices [25], [26].

Collectivités locales

Les collectivités locales expriment plusieurs enjeux autour de l'implantation de la 5G dans les territoires. La mission d'étude de la 5G de la ville de Rennes, qui a rendu ses conclusions début 2021, faisait interagir des citoyens et citoyennes

avec des élues et élus locaux [62]. Ainsi, le groupe des élues et élus locaux souligne notamment que : « *La 5G porte des opportunités en matière de création d'emploi local (nouveaux métiers, ville intelligente, intelligence artificielle, véhicules connectés, usines du futur, Orange 5G lab.) sur le territoire rennais, mais aussi des risques de destruction d'emploi* » [62]. Le déploiement de la 5G est ainsi mis en avant comme un atout à double tranchant sur le plan des opportunités socio-économiques en matière d'emploi local.

Le positionnement des collectivités locales sur les questions du déploiement de la 5G est complexe et leurs marges de manœuvre sont faibles [27]. En effet, l'attribution des fréquences est régie par l'État et le déploiement est entièrement géré par les opérateurs, avec des obligations en termes de nombre d'antennes déployées et de taux de couverture à l'échelle du territoire national (voir la section « Faut-il différencier les besoins en 5G selon les territoires ? » sur les différences de déploiement entre zones urbaines et zones rurales). Certaines communes, telle que Grenoble [92] ou Orly [93] par exemple, s'opposent au déploiement de certaines stations 5G en s'appuyant pour cela sur le code de l'urbanisme et le plan local d'urbanisme [27]. Cette opposition repose en particulier sur des arguments liés à des enjeux démocratiques, de controverse sanitaire et au principe de précaution [95]. Il est à noter que les recommandations de la Convention citoyenne pour le climat en 2020 allaient dans ce sens avec sa proposition PT12.1 « *d'instaurer un moratoire sur la mise en place de la 5G en attendant les résultats de l'évaluation de la 5G sur la santé et le climat* » [36].

Ces enjeux locaux autour de la 5G se superposent aux enjeux locaux pour le déploiement de la fibre optique, qui est également une technologie d'accès à Internet à haut débit. En effet, en 2013, l'État français a lancé le plan France très haut débit (FTHD) avec comme objectif de garantir à toutes et tous un haut débit (100 Mbps) d'ici 2025. Pour cela, les opérateurs assurent les déploiements dans les zones denses (zones urbaines), tandis que les collectivités territoriales financent les déploiements dans les zones peu denses (zones rurales où les déploiements ne seraient pas assez rentables pour les opérateurs) [43]. Le financement de cette dernière catégorie de déploiements provient ainsi principalement des régions, des départements, des intercommunalités, de l'État et de l'Union Européenne. Début 2017, la cour des comptes estimait le coût total du plan FTHD à 34,9 milliards d'euros [39]. À titre d'exemple, en Bretagne, où 90 % des territoires sont ruraux, donc non couverts par les opérateurs privés, l'investissement public s'élève à 1,7 milliard d'euros pour le plan FTHD [78]. Les retombées économiques d'un tel déploiement sont difficiles à évaluer en particulier pour les entreprises. Ainsi, une étude économétrique récente s'intéresse aux effets du plan FTHD sur la numérisation des entreprises [43]. Cette étude montre une corrélation positive entre usages numériques des entreprises et développement économique (évalué en termes de croissance économique, de création d'entreprises ou de productivité) [43]. Elle conclut notamment : « *La*

fibre bénéficie principalement aux entreprises situées dans les zones périurbaines, qui ont à la fois des incitations fortes à utiliser les outils numériques et la capacité suffisante de se les approprier. Les retombées sont en revanche quasi inexistantes dans les espaces ruraux» [43]. Cet exemple illustre que les retombées espérées d'un déploiement de réseau d'accès très haut débit ne se réalisent pas toujours et restent difficiles à évaluer. L'étude économétrique citée ci-dessus se concentre uniquement sur la numérisation des entreprises grâce au plan FTHD et n'étudie donc pas les autres retombées potentielles, telle que la réduction de la fracture numérique par exemple. Cependant, les interrogations des collectivités locales sur les déploiements de la 5G sont également à mettre en regard de leur investissement massif dans le déploiement de la fibre.

Ce positionnement des territoires vis-à-vis du déploiement de la fibre et de la 5G est très dépendant du pays. Par exemple, aux États-Unis, le déploiement de la fibre est principalement supporté par les opérateurs et le réseau fibre est peu développé hors des grandes villes [45]. En 2020, dans le cadre d'une étude géopolitique sur la neutralité du net aux États-Unis, Charlotte Escorne a mené une série d'entretiens avec différents acteurs concernés par ces déploiements (acteurs locaux, entreprises, opérateurs, etc.) [45]. Elle indique notamment : «*Nous avons interrogé des acteurs sur les enjeux de l'installation d'un réseau 5G sur un territoire comme la Californie, extrêmement consommateur en bande passante et où les sociétés de la Silicon Valley dépendent largement de l'utilisation d'Internet. Il est souvent ressorti de ces entretiens que le développement de la 5G n'était qu'un argument marketing des FAI pour justifier la volonté de supprimer la neutralité du Net et un moyen de convaincre les consommateurs du réseau qu'ils ont besoin de cette technologie, alors qu'il existe pourtant une alternative à la 5G, bien plus performante et que les industriels maîtrisent très bien : la fibre*» [45]. Les enjeux locaux identifiés dans ce cas concernent la difficulté de déploiement de la fibre, qui nécessite de la part des opérateurs une entente avec les collectivités locales, et une redistribution des revenus vers ces collectivités [45].

Le coût financier du déploiement de la fibre à large échelle demeure un enjeu majeur pour de nombreux pays, il l'est également pour la 5G. Ainsi, une étude économique réalisée en 2017 pour le territoire du Royaume-Uni montre qu'une couverture totale avec une bande passante de 10 Mbps est atteignable avec les niveaux d'investissements prévus [71]. Cependant, si un débit plus élevé est visé, les coûts deviennent prohibitifs dans certaines zones rurales. En effet, pour une bande passante de 50 Mbps, il faudrait un investissement de 6 milliards de livres sterling pour couvrir 90 % du territoire anglais et 12 milliards supplémentaires pour couvrir les 10 % restants du territoire [71]. Ceci est principalement dû aux limites de l'intégration des nouvelles fréquences de la 5G au spectre des cellules existantes dans les zones rurales. Une analyse similaire conduit à des résultats analogues sur le territoire des Pays-Bas [72].

La 5G va-t-elle induire une augmentation du trafic ?

Des facteurs multiples

Comme nous l'avons vu dans la section « Quelles sont les justifications techniques au développement de la 5G et de qui émanent-elles ? » de la première partie de ce rapport [31], les principales raisons mises en avant pour le développement de la 5G sont d'une part d'accompagner l'augmentation du trafic en repoussant les limites de capacité du réseau actuel, et d'autre part de rendre possibles de futures applications en proposant des performances nettement supérieures à celles de la 4G, en particulier en matière de débit. Ainsi, l'augmentation du trafic est non seulement facilitée par la 5G, mais elle est aussi, indirectement, un objectif de son développement puisqu'une partie des applications envisagées seront très consommatrices en données. Le développement de ces nouvelles applications et l'augmentation du trafic sont également présentés comme une nécessité économique pour les entreprises du secteur, notamment les opérateurs et les constructeurs d'équipements mobiles et de smartphones.

L'ITU (International Telecommunication Union) spécifie que la 5G devra au minimum supporter les connexions d'un million d'équipements par kilomètre carré en mode *Massive Machine Type Communications* [54]. Ces équipements, qui forment l'Internet des objets, peuvent être de natures très variées et supporter des applications très diverses. Leur densité pourra entraîner localement un grand nombre de connexions, mais il est difficile d'estimer le volume de trafic qu'ils généreront. Ce trafic viendra cependant s'ajouter au reste du trafic. Concernant le mode *Enhanced Mobile Broadband*, parmi les applications gourmandes en trafic déjà existantes, on peut citer le streaming vidéo. Ainsi, pour une vidéo 4K (3840×2160 pixels) compressée, le débit nécessaire varie entre 35 Mbps et 68 Mbps [20]. La résolution 4K pourra être utilisée pour des applications de réalité virtuelle ou augmentée. Cependant, les smartphones HD ont une résolution maximale inférieure (1920×1080 pixels) [57], nécessitant des débits de 2 à 8 Mbps pour un streaming vidéo en qualité maximale [56]. Ces débits sont déjà possibles avec une antenne 4G. Néanmoins, une antenne 4G supporte moins de connexions simultanées pour du streaming vidéo en qualité maximale qu'une antenne 5G.

D'une manière générale, évaluer précisément l'effet sur le trafic réseau de l'augmentation des débits ou de la capacité des stations est un exercice délicat. Les modèles économiques et en particulier les modèles d'abonnement associés au déploiement de chaque nouvelle génération contribuent à façonner les comportements. Par exemple, à Taïwan l'introduction de forfaits illimités a induit une forte augmentation de la consommation de données [90]. L'augmentation du trafic réseau n'est donc pas une contrainte exogène à laquelle il faudrait absolument

répondre, elle résulte d'une combinaison de facteurs incluant les nouvelles capacités et services offerts par le réseau ainsi que les modèles d'affaire qui les accompagnent. Ainsi, la compréhension des effets de la 5G sur le trafic ne saurait se limiter aux aspects purement techniques décrits ci-dessus.

L'introduction de la 4G en France

Il est difficile de prédire l'effet de l'introduction de la 5G sur le trafic mobile, cependant on peut observer l'effet produit par l'introduction de la génération mobile précédente. Les bilans trimestriels de l'ARCEP [19] permettent d'avoir accès aux volumes de données échangées sur les réseaux mobiles pour la période 2009–2021, ainsi qu'au nombre d'installations réseau et cartes SIM actives 3G et 4G (voir les figures 2 et 3).

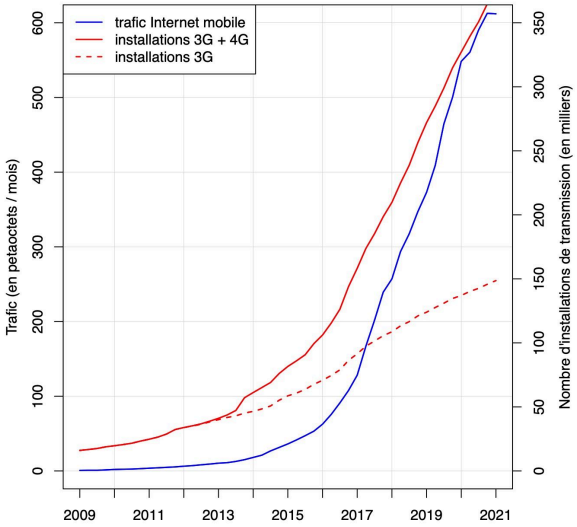


Fig. 2. Volume de données consommées sur les réseaux mobiles français et nombre d'installations 3G et 4G (obtenus à partir des rapports trimestriels de l'ARCEP et des bases de données de l'Agence Nationale des Fréquences).

On constate que lorsque les premiers abonnements 4G sont lancés début 2014, le déploiement de la 3G n'est pas terminé. En effet, depuis 2014 le nombre d'installations 3G a plus que doublé et le nombre de cartes SIM actives sur les

réseaux 3G a presque doublé. Ainsi, les déploiements de la 3G et de la 4G ne se font pas en séquence mais, pour partie, en parallèle, si bien qu'il est difficile d'isoler les effets spécifiques de la 4G sur le trafic. L'adoption de la 3G, puis de la 4G, par les utilisateurs se fait à rythme modéré, avec une croissance au plus linéaire sur la période considérée (voir la figure 3). La croissance du trafic et des infrastructures est, elle, beaucoup plus rapide : on observe une progression exponentielle jusqu'en 2017 environ, puis une progression linéaire depuis lors (voir les figures 2 et 4).

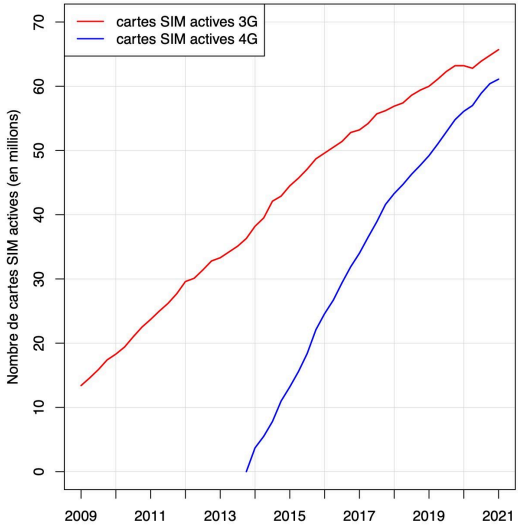


Fig. 3. Nombre de cartes SIM actives sur les réseaux français 3G et 4G (obtenus à partir des rapports trimestriels de l'ARCEP).

Nous avons confirmé ces tendances en testant différents modèles de régression : linéaire, quadratique, et exponentiel. Ces modèles ont été appliqués en distinguant deux périodes consécutives sur le jeu de données de trafic et d'installations[6], et en considérant différentes dates de séparation entre ces deux périodes. Le meilleur ajustement, c'est-à-dire minimisant l'erreur quadratique moyenne, est représenté sur la figure 4. Il indique que la croissance des infrastructures précède d'environ 9 mois celle du trafic. Cela peut suggérer que la capacité des infrastructures réseau joue un rôle dans l'augmentation du trafic, ou que les infrastructures sont déployées en prévision de cette augmentation. Ces données de trafic ne permettent pas de trancher, mais on remarque

6. Nous avons fait le choix d'exclure la période 2020–2021 correspondant à la crise Covid.

néanmoins des tendances de croissances très fortes. En effet, la consommation mensuelle moyenne de données sur les réseaux mobiles en 2011 était de 0,08 Go par carte SIM (hors cartes MtoM) [10] et elle était de 9,7 Go par carte SIM en 2021 [14], soit une multiplication par 100 du trafic par carte SIM en l'espace de 10 ans. Dans le même temps, le nombre de cartes SIM est passé de 65,2 millions en 2011 [10] à 80,7 millions en 2021 [14] (hors cartes MtoM), soit une augmentation de 24 %, cette augmentation étant plus forte que l'augmentation de la population française. Dans le même temps, les cartes MtoM (*Machine-to-Machine*) sont passées de 3,4 millions en 2011 [8] à 23,6 millions fin 2021 [14].

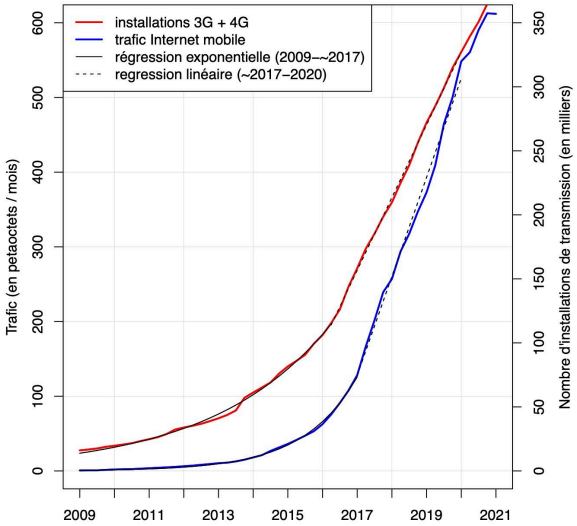


Fig. 4. Confirmation des tendances exponentielles puis linéaires à l'aide de régressions.

Une étude similaire au Royaume-Uni, exploitant les données de l'Ofcom (l'autorité en charge de la régulation des télécommunications au Royaume-Uni), montre également une forte augmentation de la consommation mensuelle moyenne de données par utilisateur de réseau mobile entre 2011 et 2016 [73] et conclut : « *The reasons behind this large increase in data demand are mainly related to the rollout of 4G LTE, allowing for higher mobile broadband speeds, driving data consumption, particularly from video and 'data-hungry' applications.* »

Les premiers déploiements de la 5G dans le monde

Hors France, il semblerait qu'on assiste à une forte augmentation du trafic dans les pays où la 5G est déployée [55], [90], [91], mais il convient d'être prudent avant d'extrapoler ces observations préliminaires en dehors des pays concernés. Une étude indique par exemple [69] qu'en 2020 les utilisateurs 5G consommaient entre 1,7 et 2,7 fois plus de données que les utilisateurs 4G (pays considérés : Corée du Sud, Royaume-Uni, Japon, États-Unis, Australie, Allemagne). Cependant, ces chiffres ne peuvent faire office de prédiction fiable car les premiers utilisateurs ne sont pas représentatifs, les gros consommateurs basculant souvent en premier sur la génération suivante de téléphonie mobile par attrait des nouvelles technologies [69]. En d'autres termes, il est souvent difficile de savoir si c'est l'adoption de la 5G qui provoque cette surconsommation, ou si au contraire c'est la surconsommation de certains utilisateurs qui les conduit à adopter la 5G (ou si c'est les deux).

Une étude récente a cependant établi un lien clair entre la vitesse du réseau et la consommation de données des utilisateurs [70]. Selon cette étude, dans le cas de la Corée du Sud, de l'Arabie Saoudite, du Royaume-Uni et des États-Unis, la différence de consommation de données entre les utilisateurs 4G et 5G ne serait pas due au profil particulier des premiers adoptants de la 5G, mais bien à un effet induit par les performances du réseau. Plus précisément, cette différence s'expliquerait principalement par l'augmentation de la qualité des flux audio et vidéo, et par l'utilisation du réseau 5G à la place des réseaux WiFi.

Faut-il différencier les besoins en 5G selon les territoires ?

La question de la fracture numérique territoriale revient de manière régulière dans le débat public depuis plus de deux décennies et notamment lorsque la technologie évolue pour offrir des débits accrus. Plusieurs rapports et plans publics mentionnent les enjeux suivants [23], [85], [24], [5] : la possibilité pour les citoyens de communiquer, de s'informer, de travailler, de consommer et de se divertir quel que soit leur lieu de résidence ; l'enjeu stratégique pour les entreprises d'améliorer leur productivité par exemple en automatisant certaines tâches ou en facilitant leurs relations avec clients et fournisseurs ; l'amélioration de la productivité des agents publics et la facilitation des démarches administratives. Les réseaux haut débit sont ainsi vus comme une condition nécessaire pour le développement économique des territoires, leur attractivité et leur cohésion [2]. Or le fonctionnement actuel du marché entre opérateurs n'a pas permis aux yeux de certaines associations de consommateurs [94] ou d'élus locaux de répondre à ces enjeux.

Dans le domaine des réseaux mobiles, l'État s'est engagé en 2018 dans un plan (le « New deal mobile ») permettant de « *résorber la fracture numérique et assurer une couverture mobile de qualité sur tout le territoire* » en accélérant le déploiement de sites 4G en zone rurale [4]. En ce qui concerne les réseaux 5G plus spécifiquement, l'attribution des licences s'est accompagnée d'une obligation de déploiement « *concomitante entre les territoires* » (25 % des sites doivent être déployés dans des zones peu denses fin 2025), d'une obligation d'augmentation des performances sur un nombre croissant de sites selon un échéancier fixé par l'ARCEP, d'obligations de couverture des axes routiers et de partage des réseaux dans les zones blanches [3]. Les pouvoirs publics reconnaissent donc l'existence d'une fracture numérique territoriale et ont tenté depuis plusieurs années d'y remédier grâce à la régulation du marché et des incitations à destination des acteurs privés.

Dans ce contexte, peut-on s'attendre à ce que la 5G accentue ou au contraire résorbe la fracture numérique territoriale ? On peut d'abord noter que la 5G nécessite de lourds investissements de la part des opérateurs. Dans une logique de rentabilité, ils ne sont donc pas naturellement poussés à déployer en priorité des sites 5G dans les zones peu denses, d'autant moins que ces sites peuvent s'avérer coûteux en termes de génie civil, d'alimentation électrique ou de connexion par le réseau d'amenée. On peut penser par conséquent que les politiques publiques visant à corriger les inégalités d'accès au réseau entre territoires, déjà mises en place pour la 4G, vont se poursuivre avec la 5G.

Ensuite, la 5G utilise différentes fréquences de 700 MHz à 3,5 GHz, puis probablement à 26 GHz. Alors que les fréquences les plus basses offrent des portées plus importantes, ce ne sont que dans les hautes fréquences que les débits seront les plus importants [37]. Cela s'explique par le fait que les bandes passantes sont plus grandes et qu'on peut utiliser plus d'antennes MIMO par station. Or le rayon des cellules à hautes fréquences est beaucoup plus réduit (quelques centaines de mètres à 3,5 GHz, pas plus de 200 m à 26 GHz, contre plusieurs kilomètres à 700 MHz). Pour une même région à couvrir, il faut environ 100 fois plus de stations à 26 GHz que de stations à 700 MHz. Il peut y avoir un facteur 10 en nombre de stations entre 3,5 GHz et 700 MHz. Bénéficier des plus hautes performances de la 5G partout sur le territoire national nécessiterait donc le déploiement d'un nombre considérable d'antennes pour lesquelles il faudrait assurer la maintenance, l'alimentation électrique et le raccordement aux réseaux fixes. À contrario, la 5G à 700 MHz offre des débits crêtes comparables à ceux de la 4G à la même fréquence [37]. En d'autres termes, à bande passante égale et nombre d'antennes MIMO égal, la 5G n'est pas beaucoup plus performante que la 4G.

Si l'égal accès de tous à un service mobile de qualité et la réduction de la fracture numérique territoriale sont des objectifs louables et sont des facteurs de cohésion sociale, il convient de s'interroger sur le niveau de performance visé

tant les coûts financiers et environnementaux semblent prohibitifs si l'on veut généraliser les performances maximales de la 5G [71]. On a en effet vu dans la section « Quel est l'inconvénient d'utiliser de nouvelles bandes de fréquences ? » de la première partie de ce rapport [31] que les plus hauts débits sont obtenus avec les fréquences les plus hautes, c'est-à-dire avec les tailles de cellules les plus petites. Offrir les plus hauts débits sur l'ensemble du territoire nécessiterait donc le déploiement d'un très grand nombre de stations de base.

Par exemple, une étude d'universitaires américains s'est penchée sur la couverture 4G de l'opérateur mobile Verizon en 2023 aux États-Unis [97]. Cette étude s'appuie sur des données ouvertes issues d'OpenCelliD [68], la couverture obtenue est illustrée par la figure 5. L'étude indique que pour obtenir une couverture identique avec des stations 5G d'une portée moyenne de 450 m (donc à haut débit), il faudrait environ 600 millions de stations 5G. Ces 600 millions de stations 5G offriraient des débits supérieurs dans les zones déjà couvertes, mais il faudrait ajouter encore d'autres stations pour couvrir les zones blanches déjà existantes. D'après OpenCelliD, il y a actuellement environ 4,9 millions de stations de base 4G tous opérateurs confondus aux États-Unis [68]. La question de la fracture numérique territoriale s'est déjà posée et se pose encore pour la 4G, mais on comprend qu'elle est accentuée avec la 5G car les cellules offrant les plus hauts débits sont aussi celles qui ont les rayons de couverture les plus petits (cf. section « Quel est l'inconvénient d'utiliser de nouvelles bandes de fréquences ? » de la première partie de ce rapport [31]).

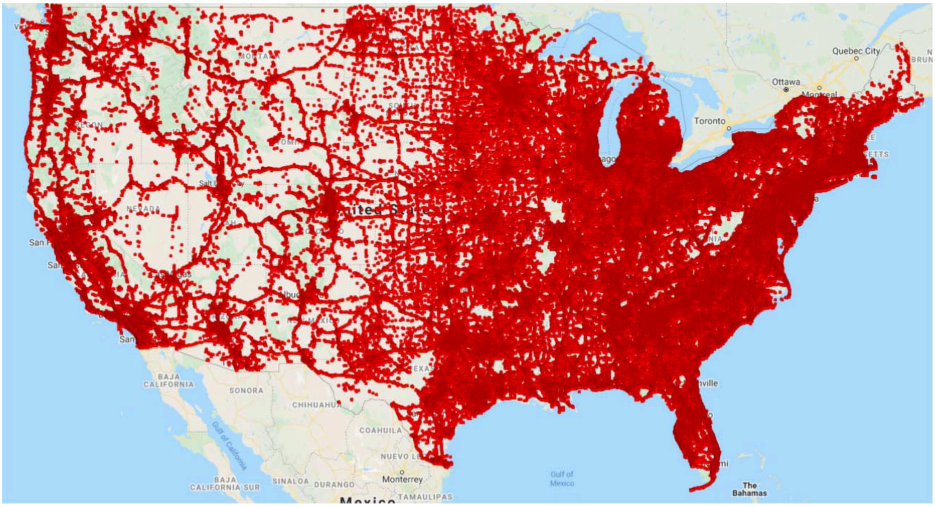


Fig. 5. Couverture 4G fournie par l'opérateur Verizon aux États-Unis d'après [96] estimée grâce aux données OpenCelliD [68].

ÉTUDE COMPARÉE DE LA 5G EN ZONE RURALE ET EN ZONE URBAINE

La figure 6 montre les débits théoriques moyens par utilisateur en fonction de la charge de la cellule dans les milieux urbain et rural pour plusieurs largeurs de bande et à différentes fréquences [38]. On considère que la fréquence à 3,5 GHz accompagnée d'une large bande de 50 ou 100 MHz est utilisée en milieu urbain. Dans les zones rurales, on suppose que la fréquence 700 MHz est utilisée avec une largeur de bande plus faible de 10, 20 ou 30 MHz. L'un des scénarios comprend une porteuse à 700 MHz avec 10 MHz de bande agrégée à une porteuse supplémentaire (SDL) à 1,4 GHz de 20 MHz, pour un total de 30 MHz. On suppose également qu'à 3,5 GHz, on utilise quatre flux MIMO simultanés contre deux à 700 MHz. Cette différence s'explique par la taille des antennes qui est beaucoup plus importante à 700 MHz.

Dans le modèle considéré, les utilisateurs arrivent de manière aléatoire dans la cellule, téléchargent un fichier de taille fixe puis quittent la cellule. Quand plusieurs utilisateurs sont en train de télécharger, ils se partagent les ressources de la station. On définit alors la charge de la cellule comme la proportion de ressources utilisée par la station pour servir les utilisateurs. Si la charge est de x ($0 \leq x \leq 1$), alors la station émet pendant $x \times 100\,\%$ du temps et n'émet pas pendant $(100 - x) \times 100\,\%$ du temps. Quand la charge est très faible, un utilisateur qui arrive dans la cellule est très certainement seul, il bénéficie alors de toutes les ressources de la station pendant son téléchargement et obtient ainsi un débit élevé. Quand la charge est élevée, les utilisateurs actifs doivent se partager les ressources, leur débit est donc faible. À la limite, quand la charge est de 100 %, le débit par utilisateur tend vers 0 et il faut un temps infini pour télécharger le fichier.

On remarque que, pour une charge donnée, plus la bande est large (par exemple 100 MHz au lieu de 10 MHz), plus le débit par utilisateur est élevé. Ceci explique qu'à charge équivalente, les cellules en zone rurale offrent des débits beaucoup plus faibles. Ce phénomène est renforcé par le plus grand nombre d'antennes MIMO à 3,5 GHz.

Cependant, on note qu'une cellule très faiblement chargée en zone rurale est susceptible de fournir un débit moyen par utilisateur plus élevé qu'une cellule très chargée en zone urbaine et ce, malgré les différences de largeur de bande et d'antennes MIMO. Par exemple, le débit d'un utilisateur urbain dans une cellule avec largeur de bande de 50 MHz chargée à 80 % est d'environ 30 Mbps, alors que le débit d'un utilisateur rural dans une cellule avec largeur de bande 30 MHz chargée à 5 % est d'environ 50 Mbps. Comme nous ne connaissons pas la distribution actuelle des charges des cellules car ces données ne sont pas publiques, il n'est pas possible de conclure sur les différences de qualité de service entre les zones rurales et les zones urbaines. En pratique cependant, les débits restent nettement inférieurs en zone rurale car les cellules sont plus grandes (ce qui augmente la charge et réduit la qualité du canal à cause d'un affaiblissement de parcours plus important) et que les allocations en 2024 font la part belle aux hautes fréquences (70 ou 90 MHz à 3,5 GHz contre 5 à 10 MHz à 700 MHz).

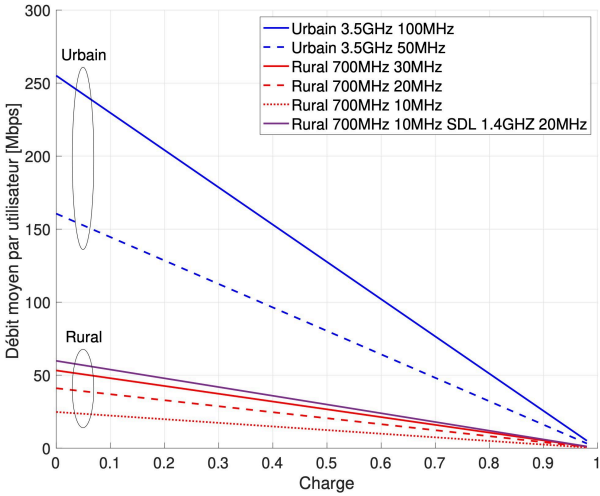


Fig. 6. Débits moyens des utilisateurs en zones urbaines et rurales en fonction de la charge de la cellule.

En conclusion, l'accélération des déploiements 4G et 5G en zone rurale devrait permettre à ces territoires un meilleur accès à l'ensemble des services numériques, mais il est dans le même temps possible que les inégalités territoriales s'accentuent avec l'introduction de la technologie 5G puisque la qualité des connexions en zones urbaines devrait progresser encore plus rapidement.

Remerciements

Les auteurs tiennent à remercier (par ordre alphabétique) : Louis Golard, Matthieu Latapy, Nicolas Montavont et Loutfi Nuaymi, pour leur relecture attentive et leurs commentaires constructifs.

Annexe A. Volume moyen mensuel par carte SIM sur les réseaux mobiles français

Le tableau suivant reprend les valeurs utilisées pour la figure 1 pour les années 2009 à 2022 et indique le rapport annuel de l'ARCEP d'où vient chaque valeur.

Année	Vol. Go	Source
2009	0,02	ARCEP [8], page 71
2010	0,043	ARCEP [8], page 71
2011	0,076	ARCEP [8], page 71
2012	0,121	ARCEP [9], page 70
2013	0,192	ARCEP [9], page 70
2014	0,367	ARCEP [9], page 70
2015	0,662	ARCEP [9], page 70
2016	1,243	ARCEP [9], page 70
2017	2,7	ARCEP [15], page 62
2018	4,3	ARCEP [15], page 62
2019	6,1	ARCEP [15], page 62
2020	8,2	ARCEP [15], page 62
2021	9,7	ARCEP [15], page 62
2022	12,1	ARCEP [16], page 61
2023	14,3	ARCEP [18], page 43

Table 1. Volumes moyens par carte SIM (France).

Références

[1] UE 2015. Règlement (UE) 2015/2120) du parlement européen et du conseil du 25 novembre 2015 établissant des mesures relatives à l'accès à un internet ouvert et modifiant la directive 2002/22/CE concernant le service universel et les droits des utilisateurs au regard des réseaux et services de communications électroniques et le règlement (UE) nº 531/2012 concernant l'itinérance sur les réseaux publics de communications mobiles à l'intérieur de l'Union.

[2] Sénat 2019. Consultation Publique sur l'Attribution des fréquences 5G. Position de la commission de l'aménagement du territoire et du développement durable.

[3] ARCEP. Décision nº 2019-1386 du 21 novembre 2019 de l'Autorité de régulation des communications électroniques, des postes et de la distribution de la presse proposant au ministre chargé des communications électroniques les modalités et les conditions d'attribution d'autorisations d'utilisation de fréquences dans la bande 3,4 - 3,8 GHz en France métropolitaine pour établir et exploiter un réseau radioélectrique mobile ouvert au public.

[4] New Deal mobile : plus de 600 nouveaux sites 4G en zone rurale. Communiqué du Ministère de l'économie, des finances et de la souveraineté industrielle et numérique. `https://www.economie.gouv.fr/new-deal-mobile-plus-600-nouveaux-sites-4g-zone-rurale`, 2022.

[5] Le plan France Très Haut Débit (PFTHD). https://www.arcep.fr/demarches-et-services/collectivites/le-plan-france-tres-haut-debit-pfthd.html, 2022.

[6] 3GPP. 2019. *Study on physical layer enhancements for NR ultra-reliable and low latency case (URLLC)*. 3rd Generation Partnership Project (3GPP). Retrieved from http://www.3gpp.org/DynaReport/38824.htm.

[7] 3HK. 2024. 5G Stock Pro Monthly Plan. https://web.three.com.hk/plans/finsim/index-en.html.

[8] ARCEP. 2015. Observatoire annuel du marché des communications électroniques en France - Année 2013 - Résultats définitifs. https://www.arcep.fr/fileadmin/reprise/observatoire/marc-an2013/obs-annuel-2013-def-080115.pdf.

[9] ARCEP. 2017. Les services de communications électroniques en France - Année 2016 - Résultats définitifs. https://www.arcep.fr/fileadmin/cru-1677573101/reprise/observatoire/march-an2016/obs-marches-2016-def-211217.pdf.

[10] ARCEP. 2020. Marché des communications électroniques en France - Année 2019 - Résultats définitifs. https://www.arcep.fr/cartes-et-donnees/nos-publications-chiffrees/observatoire-des-marches-des-communications-electroniques-en-france/marche-des-communications-electroniques-en-france-annee-2019-resultats-definitifs.html.

[11] ARCEP. 2020. 5G - L'Arcep crée un observatoire des déploiements 5G. https://www.arcep.fr/fileadmin/cru-1677573101/user_upload/55-20-version-francaise.pdf.

[12] ARCEP. 2022. Qualité des services mobiles. https://www.arcep.fr/fileadmin/cru-1664269931/user_upload/53-22-version-francaise.pdf.

[13] ARCEP. 2022. Observatoire du déploiement 5G. https://www.arcep.fr/fileadmin/cru-1669014821/user_upload/observatoire/5G-deploiements/septembre-2022/Observatoire5G_France_29092022.pdf.

[14] ARCEP. 2022. Marché des communications électroniques en France - Les chiffres au 1er trimestre 2022. https://www.arcep.fr/cartes-et-donnees/nos-publications-chiffrees/observatoire-des-marches-des-communications-electroniques-en-france/t1-2022.html.

[15] ARCEP. 2022. Les services de communications électroniques en France - Année 2021 - Résultats définitifs. https://www.arcep.fr/cartes-et-donnees/nos-publications-chiffrees/observatoire-des-marches-des-communications-electroniques-en-france/marche-des-communications-electroniques-en-france-annee-2021-resultats-definitifs.html.

[16] ARCEP. 2023. Les services de communications électroniques en France - Année 2022 - Résultats définitifs. https://www.arcep.fr/cartes-et-donnees/nos-publications-chiffrees/observatoire-des-marches-des-communications-electroniques-en-france/marche-communications-electroniques-france-2022-resultats-definitifs.html.

[17] ARCEP. 2023. L'état d'internet en france. https://www.arcep.fr/uploads/tx_gspublication/RA-2023-TOME3_etat-internet-france_juillet2023.pdf.

[18] ARCEP. 2024. Les services de communications électroniques en France - Année 2023 - Résultats provisoires. https://www.arcep.fr/cartes-et-donnees/nos-publications-chiffrees/observatoire-des-marches-des-communications-electroniques-en-france/marche-communications-electroniques-france-2023-resultats-provisoires.html.

[19] ARCEP. Observatoire des marchés des communications électroniques en France - Historique de l'observatoire. https://www.arcep.fr/cartes-et-donnees/nos-publications-chiffrees/observatoire-des-marches-des-communications-electroniques-en-france/historique-de-lobservatoire.html.

[20] Ghufran Baig, Jian He, Mubashir Adnan Qureshi, Lili Qiu, Guohai Chen, Peng Chen, and Yinliang Hu. 2019. Jigsaw: Robust Live 4K Video Streaming. In *Annual international conference on mobile computing and networking*. Retrieved from https://doi.org/10.1145/3300061.3300127.

[21] Martina Barbi, Alejandro Antón Ruiz, Arturo Mrozowski Handzel, Saúl Inca, David Garcia-Roger, and Jose F. Monserrat. 2022. Simulation-based Digital Twin for 5G Connected Automated and Autonomous Vehicles. In *Joint european conference on networks and communications & 6G summit (EuCNC/6G summit)*, 273–278. https://doi.org/10.1109/EuCNC/6GSummit54941.2022.9815732.

[22] Luis Bastos, Germano Capela, Alper Koprulu, and Gerard Elzinga. 2021. Potential of 5G technologies for military application. In *International conference on military communication and information systems (ICMCIS)*, 1–8. https://doi.org/10.1109/ICMCIS52405.2021.9486402.

[23] Claude Belot. 2005. Haut débit et territoires : enjeu de couverture, enjeu de concurrence. Rapport d'information nº 443 (2004-2005), Sénat, https://www.senat.fr/rap/r04-443/r04-443_mono.html.

[24] Éric Besson. 2008. France numérique 2012 - Plan de développement de l'économie numérique. https://www.vie-publique.fr/rapport/30143-france-numerique-2012-plan-de-developpement-economie-numerique.

[25] Colin Blackman and Simon Forge. 2016. European Leadership in 5G. Rapport de la Commission de l'industrie, de la recherche et de l'énergie (ITRE) du Parlement européen. https://www.europarl.europa.eu/RegData/etudes/IDAN/2016/595337/IPOL_IDA(2016)595337_EN.pdf.

[26] Knut Blind and Crispin Niebel. 2022. 5G roll-out failures addressed by innovation policies in the EU. *Technological Forecasting and Social Change* 180:121673. https://doi.org/10.1016/j.techfore.2022.121673.

[27] Edouard Bouté. 2020. Débats 5G – Quels apports de la recherche?. Rapport Centre Internet et Société - CNRS; Fondation Internet Nouvelle Génération, \url{https://shs.-hal.science/halshs-03595461/document.

[28] Victor Boutin, Alexis Hannart, Abderrahim Essaidi, and Brunilde Sansò. 2021. Offloading autonomous vehicle machine learning algorithms to the 5G edge: A proof of concept implementation. In *IEEE 4th 5G world forum (5GWF)*, 269–274. https://doi.org/10.1109/5GWF52925.2021.00054.

[29] Stefana Broadbent, Florian Forestier, Mehdi Khamassi, and Célia Zolynski. 2024. *Pour une nouvelle culture de l'attention*. Odile Jacob.

[30] Pierre-Antoine Chardel and Olivier Frérot. 2022. Controverse sur la 5G. *Études* 1 : 57–66.

[31] Philippe Ciblat, Jacques Combaz, Marceau Coupechoux, Kevin Marquet, and Anne-Cécile Orgerie. 2022. *Effets environnementaux de la 5G – partie 1*. EcoInfo. Retrieved from https://hal.science/hal-03810501/file/EcoInfo-5G-partie1.pdf.

[32] Cisco. 2020. Cisco Annual Internet Report 2018-2023. White paper. https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf.

[33] Commission des participations et des transferts. 2019. Avis nº 2019 – A. – 8 du 22 novembre 2019 relatif à l'évaluation de lots de fréquences hertziennes dans la bande de 3,4-3,8 GHz.

[34] Commission Européenne. 2019. Commission Recommendation of 26 March 2019 on Cybersecurity of 5G networks.

[35] Commission supérieure du numérique et des postes. 2022. Avis nº 2022-05 du 27 Juillet 2022 portant recommandations sur le financement et le modèle économique des infrastructures et des réseaux de télécommunications.

[36] Convention citoyenne pour le climat. 2021. Les propositions – rapport final. https://propositions.conventioncitoyennepourleclimat.fr/pdf/ccc-rapport-final.pdf.

[37] Marceau Coupechoux. 2020. Les débits en 5G : mythes et réalité. Présentation à l'ARCEP, https://www.arcep.fr/fileadmin/user_upload/grands_dossiers/5G/debits-5G-mythes-realites.pdf.

[38] Marceau Coupechoux. 2021. *Performances 5G : étude comparée en zones rurales et urbaines*. Telecom Paristech. Retrieved from https://hal.telecom-paris.fr/hal-03367918.

[39] Cour des comptes. 2017. Les réseaux fixes de haut et très haut débit : un premier bilan. Rapport public thématique.

[40] Giovanna Culot, Guido Nassimbeni, Guido Orzes, and Marco Sartor. 2020. Behind the definition of Industry 4.0: Analysis and open questions. *International Journal of Production Economics* 226:107617. https://doi.org/10.1016/j.ijpe.2020.107617.

[41] Martin de Jong, Simon Joss, Daan Schraven, Changjie Zhan, and Margot Weijnen. 2015. Sustainable–smart–resilient–low carbon–eco–knowledge cities; making sense of a multitude of concepts promoting sustainable urbanization. *Journal of Cleaner Production* 109:25–38. https://doi.org/10.1016/j.jclepro.2015.02.004.

[42] Chen Dongxu. 2020. 5G Power: Creating a green grid that slashes costs, emissions & energy use. Huawei, https://www.huawei.com/en/technology-insights/publications/huawei-tech/89/5g-power-green-grid-slashes-costs-emissions-energy-use.

[43] Chloé Duvivier, Laurent Bergé, and Florian Léon. 2024. Le déploiement du très haut débit a-t-il favorisé la numérisation des entreprises? Une évaluation du plan France Très Haut Débit. *Revue économique* 75, 2:301–352.

[44] Ericsson. 2020. *Addressing efficiency and sustanability at the Livorno port with 5G*. Ericsson. Rapport technique.

[45] Charlotte Escorne. 2020. Les enjeux de la neutralité du Net aux États-Unis. *Hérodote* 2-3, 177-178:215–234.

[46] Charoltte Escorne. 2022. Le déploiement de la 5G en Afrique : Vers de nouvelles dynamiques territoriales et relations entre les Etats africains et leurs partenaires?. Rapport. https://hal.science/hal-04391464.

[47] Federal Communications Commission. 2024. Safeguarding and Securing the Open Internet – Restoring Internet Freedom. Ordonnances 23-320 et 17-108, https://www.fcc.gov/document/fcc-restores-net-neutrality-0.

[48] Franceinfo. 2020. 5G : les ONG environnementales dénoncent un renoncement d'Emmanuel Macron sur la Convention climat.

[49] Amitava Ghosh, Andreas Maeder, Matthew Baker, and Devaki Chandramouli. 2020. 5G evolution: a view on 5G cellular technology beyond 3GPP Release 15.

[50] Google. 2024. Google Fiber. https://fiber.google.com.

[51] Google. 2024. Pixel, le seul téléphone conçu par Google. https://store.google.com/fr/category/phones?hl=fr.

[52] GSC. 2019. *C-band is critical for satellite services*. GSC. Rapport technique.

[53] GSMA. 2020. 5G Synchronisation TDD Questions et Réponses – Recommandations pour la coexistence des réseaux TDD dans la gamme de fréquence 3,5 GHz.

[54] ITU. 2017. *ITU-R M.2410-0*. ITU. https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-M.2410-2017-PDF-E.pdf.

[55] Frederik Jungerman. 2020. Nine Months with 5G: 4.7 Million Subs – Each Using 27 GB per Month. https://tefficient.com/nine-months-with-5g-4-7-million-subs-each-using-27-gb-per-month.

[56] Dilip Kumar Krishnappa, Divyashri Bhat, and Michael Zink. 2013. DASHing YouTube: An analysis of using DASH in YouTube video service. In *Annual IEEE conference on local computer networks*, 407–415. https://doi.org/10.1109/LCN.2013.6761273.

[57] Rohit Kumar, Pedro Assuncao, Lino Ferreira, and Antonio Navarro. 2018. Retargeting UHD 4k Video for Smartphones. In *IEEE international conference on consumer electronics (ICCE)*. https://doi.org/10.1109/ICCE-Berlin.2018.8576242.

[58] Le Monde. 2020. Les enchères de la 5G rapportent 2,786 milliards d'euros à l'Etat, Orange obtient la plus grande part des fréquences en jeu.

[59] Jung Hoon Lee, Robert Phaal, and Sang-Ho Lee. 2013. An integrated service-device-technology roadmap for smart city development. *Technological Forecasting and Social Change* 80, 2:286–306. https://doi.org/10.1016/j.techfore.2012.09.020.

[60] William Lehr, Fabian Queder, and Justus Haucap. 2021. 5G: A new future for Mobile Network Operators, or not? *Telecommunications Policy* 45, 3:102086. https://doi.org/10.1016/j.telpol.2020.102086.

[61] Shancang Li, Li Da Xu, and Shanshan Zhao. 2018. 5G Internet of Things: A survey. *Journal of Industrial Information Integration* 10:1–9. https://doi.org/10.1016/j.jii.2018.01.005.

[62] Mission d'étude de la 5G de la ville de Rennes. 2021. La 5G à Rennes : le débat – rapport final. http://www.presse.metropole.rennes.fr/upload/espace/1/pj/13855_10723_Mission_5G_rapport_final.pdf.

[63] Bengt G Mölleryd, Jan Markendahl, Jan Werding, and Östen Mäkitalo. 2010. Decoupling of revenues and traffic - is there a revenue gap for mobile broadband? In *2010 9th conference of telecommunication, media and internet*, 1–7. https://doi.org/10.1109/CTTE.2010.5557696.

[64] Nokia. 2016. *5G network energy efficiency: Massive capacity boost with flat energy consumption*. Nokia. https://onestore.nokia.com/asset/200876.

[65] Observatoire de la filière 5G. 2021. Observatoire de la filière 5G en France en 2021. https://www.fftelecoms.org/5g/observatoire-de-la-filiere-5g-en-france.

[66] Ofcom. 2021. Award of the 700 MHz and 3.6-3.8 GHz spectrum bands – Notice under regulation 121 of the Wireless Telegraphy (Licence Award) Regulations 2020.

[67] Ooredoo. 2024. ProPing 5G service. https://www.ooredoo.com.kw/portal/en/proping_2_0.

[68] OpenCelliD. 2023. OpenCelliD open data. https://opencellid.org.

[69] OpenSignal. 2020. 5G users on average consume up to 2.7x more mobile data compared to 4G users. https://www.opensignal.com/2020/10/21/5g-users-on-average-consume-up-to-27x-more-mobile-data-compared-to-4g-users.

[70] OpenSignal. 2021. Quantifying the impact of 5G and Covid-19 on mobile data consumption. https://www.opensignal.com/reports/2021/06/quantifying-the-impact-of-5G-and-covid-19-on-mobile-data-consumption.

[71] Edward J. Oughton and Zoraida Frias. 2018. The cost, coverage and rollout implications of 5G infrastructure in Britain. *Telecommunications Policy* 42, 8:636–652.

[72] Edward J. Oughton, Zoraida Frias, Sietse van der Gaast, and Rudolf van der Berg. 2019. Assessing the capacity, coverage and cost of 5G infrastructure strategies: Analysis of the Netherlands. *Telematics and Informatics* 37:50–69.

[73] Edward Oughton, Zoraida Frias, Tom Russell, Douglas Sicker, and David D. Cleevely. 2018. Towards 5G: Scenario-based assessment of the future supply and demand for mobile telecommunications infrastructure. *Technological Forecasting and Social Change* 133:141–155.

[74] Sreekrishna Pandi, Frank H. P. Fitzek, Christopher Lehmann, David Nophut, Domokos Kiss, Viktor Kovacs, Akos Nagy, Gabor Csorvasi, Miklos Toth, Tamas Rajacsis, Hassan Charaf, and Rainer Liebhart. 2016. Joint Design of Communication and Control for Connected Cars in 5G Communication Systems. In *IEEE globecom workshops (GC wkshps)*, 1–7. https://doi.org/10.1109/GLOCOMW.2016.7848940.

[75] Parlement Européen. 2019. Menaces pour la sécurité liées à la présence technologique croissante de la Chine dans l'Union et les actions possibles à l'échelle de l'UE pour les réduire. https://www.europarl.europa.eu/doceo/document/TA-8-2019-0156_FR.pdf.

[76] Roxana Radu and Cedric Amon. 2021. The governance of 5G infrastructure: between path dependency and risk-based approaches. *Journal of Cybersecurity* 7, 1. https://doi.org/10.1093/cybsec/tyab017.

[77] Sriganesh Rao and Ramjee Prasad. 2018. Impact of 5G Technologies on Industry 4.0. *Wireless Personal Communications* 100:1–15. https://doi.org/10.1007/s11277-018-5615-7.

[78] Région Bretagne. 2020. La fibre pour tous en Bretagne - 2e édition. https://www.bretagne.bzh/actions/grands-projets/fibre-pour-tous/.

[79] Fabien Roussel. 2020. Question à l'Assemblée Nationale pour un moratoire sur le déploiement de la 5G. https://questions.assemblee-nationale.fr/q15/15-32100QE.htm.

[80] Gauthier Roussilhe. 2020. La controverse de la 5G. https://gauthierroussilhe.com/pdf/5G-Juillet2020.pdf.

[81] Samsung. 2015. *5G Vision*. Samsung. https://images.samsung.com/is/content/samsung/p5/global/business/networks/insights/white-paper/5g-vision/global-networks-insight-samsung-5g-vision-2.pdf.

[82] Sandvine. 2023. The Global Internet Phenomena Report. Rapport annuel, https://www.sandvine.com/hubfs/Sandvine_Redesign_2019/Downloads/2023/reports/Sandvine%20GIPR%202023.pdf.

[83] Philipp Schulz, Maximilian Matthe, Henrik Klessig, Meryem Simsek, Gerhard Fettweis, Junaid Ansari, Shehzad Ali Ashraf, Bjoern Almeroth, Jens Voigt, Ines Riedel, Andre Puschmann, Andreas Mitschele-Thiel, Michael Muller, Thomas Elste, and Marcus Windisch. 2017. Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture. *IEEE Communications Magazine* 55, 2:70–78. https://doi.org/10.1109/MCOM.2017.1600435CM.

[84] Enas Selem, Mohammed Fatehy, and Sherine M. Abd El-Kader. 2019. E-Health applications over 5G networks: challenges and state of the art. In *2019 international conference on advanced control circuits and systems (ACCS)*, 111–118. https://doi.org/10.1109/ACCS-PEIT48329.2019.9062841.

[85] Sénat. 2011. Développement du très haut débit dans les territoires ruraux. http://www.senat.fr/questions/base/2011/qSEQ111221565.html.

[86] Jasmeet Singh Sethi, Yalu Fan, Cristina Oprea, and Gunnar Ehrnborg. 2023. 5G value: Turning performance into loyalty – Exploring what drives 5G network satisfaction and user retention.

[87] Kinza Shafique, Bilal A. Khawaja, Farah Sabir, Sameer Qazi, and Muhammad Mustaqim. 2020. Internet of things (IoT) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5G-IoT scenarios. *IEEE Access* 8:23022–23040. https://doi.org/10.1109/ACCESS.2020.2970118.

[88] Gurfarmaan Singh, Robert Casson, and WengOnn Chan. 2021. The potential impact of 5G telecommunication technology on ophthalmology. *Eye* 35:1859–1868.

[89] Yu Tang, Sathian Dananjayan, Chaojun Hou, Qiwei Guo, Shaoming Luo, and Yong He. 2021. A survey on the 5G network and its impact on agriculture: Challenges and opportunities. *Computers and Electronics in Agriculture* 180:1–15. https://doi.org/10.1016/j.compag.2020.105895.

[90] Tefficient. 2019. *Mobile data consumption continues to grow – a majority of operators now rewarded with ARPU*. https://tefficient.com/wp-content/uploads/2019/09/tefficient-industry-analysis-3-2019-mobile-data-usage-and-revenue-1H-2019-per-operator-5-Sep.pdf.

[91] Tefficient. 2020. *The gig economy: 58% more gigabytes generated 0.4% more revenue 5G monetisation lessons from China and Korea*. https://tefficient.com/wp-content/uploads/2020/04/tefficient-industry-analysis-1-2020-mobile-data-usage-and-revenue-FY-2019-per-operator.pdf.

[92] Tribunal administratif de Grenoble. 22 février 2023. Affaire 2300642. https://justice.pappers.fr/decision/ee58163140b5ecbac6c93ad5fd39be7e0e6b57cf.

[93] Tribunal administratif de Melun. 10 juillet 2023. Affaire 2306181. https://justice.pappers.fr/decision/cb262f476fc20ebcc170b07bf40550b61cce7f21.

[94] UFC-Que Choisir. 2022. Les réseaux mobiles les plus performants. https://www.que-choisir.org/decryptage-telephonie-mobile-les-reseaux-mobiles-les-plus-performants-n98164/.

[95] Aline Vignon-Barrault. 2020. Les juges, le principe de précaution et les antennes-relais. *Archives de philosophie du droit* 62:255–269.

[96] Tim Wu. 2016. *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. Knopf.

[97] Z. Zhang, L. Scola, and A. Schulman. 2023. Investigating the Sustainability of the 5G Base Station Overhaul in the United States. In *International conference on ICT for sustainability (ICT4S)*, 53–62.

[98] J. Zheng, Y. Wang, J. Zhang, and others. 2020. 5G ultra-remote robot-assisted laparoscopic surgery in China. *Surgical Endoscopy* 34:5172–5180. https://doi.org/10.1007/s00464-020-07823-x.

Études, conceptions et applications des interfaces cerveau-ordinateur

Fabien Lotte[1]

Au moment d'écrire ces lignes, l'électro-encéphalographie, une technique mesurant l'activité cérébrale sous forme de microcourants électriques à l'aide d'électrodes posées sur le cuir chevelu, fête ses 100 ans. Le premier électro-encéphalogramme (EEG) a en effet été mesuré en 1924 par le psychiatre allemand Hans Berger [19]. l'EEG est utilisé depuis ce temps notamment pour effectuer des diagnostics médicaux dans le domaine clinique, ou pour comprendre le fonctionnement cérébral en neurosciences [34].

Plus récemment, l'EEG est aussi utilisé pour concevoir ce qu'on appelle des interfaces cerveau-ordinateur ou BCI (de l'anglais *brain-computer interface*) [11, 48]. Une BCI est un système traduisant une mesure de l'activité cérébrale (typiquement l'EEG) en une commande ou un message pour une application interactive. Une BCI permet par exemple de déplacer une balle dans un jeu vidéo vers la gauche ou vers la droite, en imaginant des mouvements de la main gauche ou droite respectivement (cf. figure 1), ces mouvements imaginés étant reconnus dans les signaux EEG de l'utilisateur à l'aide de technique d'apprentissage artificiel [8].

Les BCI sont prometteuses pour de nombreuses applications [12], notamment pour des applications de communication et de contrôle pour des personnes gravement paralysées, pour leur permettre de sélectionner des lettres dans un éditeur de texte [6] ou bien de déplacer un fauteuil roulant vers la gauche ou vers la droite en imaginant des mouvements de la main gauche ou droite [30].

1. Centre Inria de l'université de Bordeaux, LaBRI.

DOI:10.48556/SIF.1024.24.129

Elles peuvent aussi servir pour le grand public et les personnes en bonne santé, par exemple pour contrôler des jeux vidéos ou applications de réalité virtuelle, e.g., en imaginant des mouvements des pieds pour avancer dans un environnement virtuel [29]. Encore plus récemment, les BCI se sont montrées prometteuses pour la rééducation motrice après accident vasculaire cérébral (AVC — cf. section « Applications des BCI » plus loin dans cet article) [35], ou encore pour concevoir ce qu'on appelle des technologies neuroadaptives. Ces dernières sont des systèmes qui vont estimer l'état mental de leurs utilisateurs en continu, à partir de leurs signaux cérébraux, de façon à adapter l'interaction humain-machine en fonction [53]. Par exemple, elles peuvent estimer le niveau d'attention ou d'effort mental d'un pilote d'avion de façon à adapter la quantité d'information affichée dans le cockpit en conséquence, afin d'éviter une surcharge attentionnelle [4].

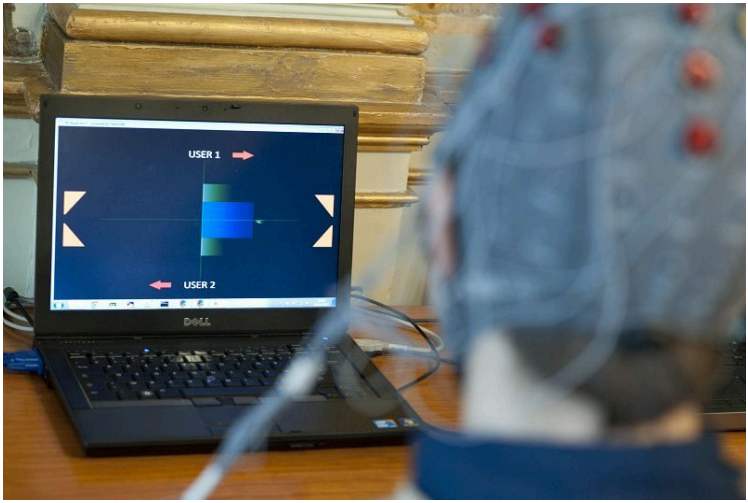
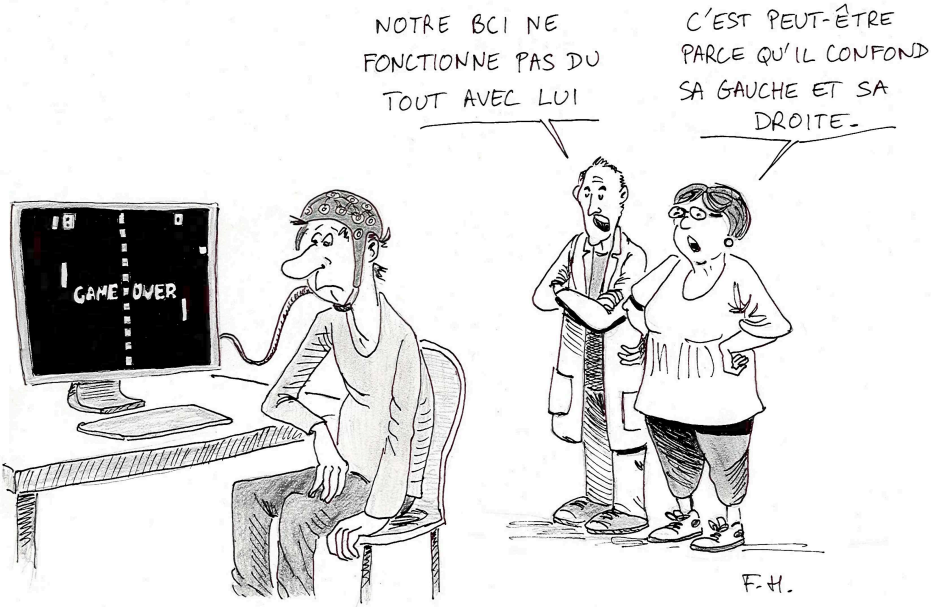


Fig. 1. Un exemple de BCI permettant à un utilisateur (dont on mesure l'activité cérébrale par EEG) de déplacer la balle sur l'écran vers la gauche ou vers la droite, simplement en imaginant des mouvements de la main gauche ou droite [8] (© Inria / Photo G. Maisonneuve).

Les BCI sont ainsi très prometteuses. Malheureusement, elles souffrent aussi d'un certain nombre de problèmes et de limitations. En particulier, même si les BCI fonctionnent, elles ne fonctionnent pas bien. Elles souffrent en effet d'un taux d'erreurs élevé, et se trompent relativement souvent de commande mentale reconnue. Par exemple, pour distinguer un mouvement imaginé de la main gauche d'un mouvement imaginé de la main droite, il est estimé que les

systèmes actuels se trompent environ 1 fois sur 4 ou 5 (par exemple, la BCI reconnaît un mouvement imaginé de la main gauche alors que l'utilisateur a imaginé un mouvement de la main droite), ce qui est très élevé [7]. On estime également qu'entre 10 % à 30 % des utilisateurs de BCI ne peuvent pas utiliser les BCI, car celles-ci ne peuvent pas reconnaître leurs commandes mentales [1]. Enfin, les BCI souffrent d'une grande variabilité en terme de performances, à la fois entre les utilisateurs mais aussi au cours du temps pour un même utilisateur [42]. Certains utilisateurs vont pouvoir utiliser la BCI quasi parfaitement (toutes leurs commandes mentales seront bien reconnues) tandis que d'autres ne pourront pas du tout les utiliser, avec tout le spectre de performances possibles entre ces deux extrêmes. Mais, même pour un unique utilisateur, les performances peuvent varier considérablement, avec un utilisateur pouvant très bien contrôler la BCI un jour donné, mais plus le lendemain, en utilisant pourtant la même BCI. Ainsi les BCI ne sont pas encore assez fiables actuellement, et par conséquent, pas encore largement appliquées en pratique, en dehors des laboratoires. Ainsi, un des défis majeurs est de concevoir des BCI qui soient fiables [15].

La suite de cet article propose un aperçu — forcément incomplet — de nos recherches sur les BCI — en commençant par présenter nos travaux de



recherche fondamentale visant précisément à rendre les BCI plus fiables. Pour cela cet article s'intéresse tout d'abord aux aspects machine des BCI (section « Apprentissage machine dans les BCI »), puis aux aspects humains et utilisateurs de ces technologies (section « Apprentissage humain dans les BCI »). Il présente ensuite certains de nos travaux applicatifs des BCI, afin de rendre les BCI utiles (section « Applications des BCI »), et se termine par une conclusion sur ces travaux et quelques perspectives dans ce domaine en section « Conclusion ».

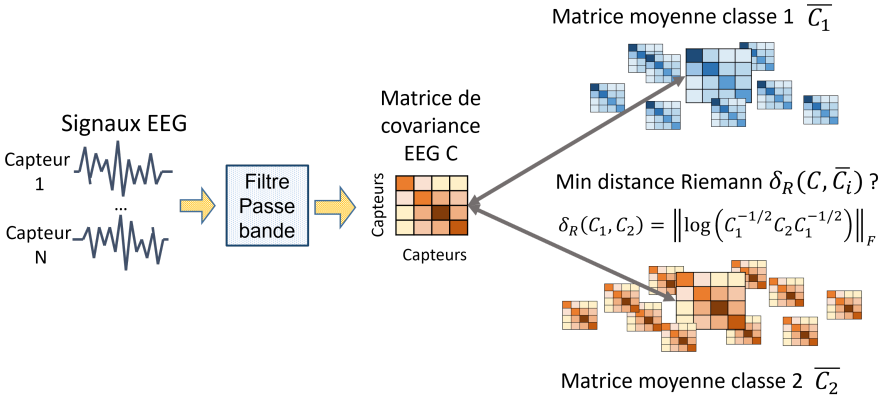


Fig. 2. Illustration du processus de classification riemannienne de signaux EEG : les signaux EEG sont d'abord filtrés dans une bande de fréquence spécifique, avant de calculer leur matrice de covariance. Cette matrice de covariance est ensuite utilisée pour classifier les signaux EEG, en assignant à la matrice de covariance la classe dont la matrice de covariance moyenne (sur les données d'apprentissage) est la plus proche, mesurée avec une distance Riemannienne [52,13].

Apprentissage machine dans les BCI

Une première façon de rendre les BCI plus fiables, est d'explorer la conception d'algorithmes de classification des signaux EEG qui soient plus robustes et plus précis, de façon à mieux reconnaître les états mentaux des utilisateurs dans leurs signaux EEG [15]. Ainsi, une part significative de la recherche sur les BCI porte sur la conception et l'étude d'algorithmes d'apprentissage artificiel et d'intelligence artificielle pour la classification de signaux EEG [25, 24]. Une part importante de nos travaux s'est donc naturellement portée sur ce sujet également (cf., par exemple, [54, 28, 27]). Dans la suite de cet article, nous nous intéresserons plus particulièrement à nos travaux plus récents sur le sujet, portant sur les classifieurs de signaux EEG dits « riemanniens ».

Les classifieurs riemanniens (CR) — qui ont été introduits dans les BCI par Alexandre Barachant et ses collègues — sont des classifieurs qui représentent les signaux EEG sous forme de matrices symétriques définies positives — généralement des matrices de covariance — et qui manipulent et classifient ces matrices avec une géométrie particulière : la géométrie riemannienne [52, 13]. Un exemple typique de CR est le « minimum distance to the riemannian mean » (MDRM). Pour l'utiliser, les signaux EEG sont tout d'abord classiquement filtrés par un passe-bande dans une banque de fréquence d'intérêt (par exemple dans les fréquences EEG sensorimotrices α (8–12 Hz) et β (16–24 Hz) pour détecter des mouvements imaginés). Ensuite, la matrice de covariance de ces signaux EEG est calculée. Cette matrice possède autant de lignes et de colonnes que le nombre d'électrodes EEG utilisées. Elle contient en diagonale la variance du signal mesuré par chaque électrode (qui est équivalente à la puissance du signal EEG dans la bande de fréquences dans laquelle il est filtré). Hors diagonale, elle contient la covariance entre les signaux de chaque paire d'électrodes, indiquant donc si ces signaux varient simultanément. La géométrie riemannienne permet de manipuler directement ces matrices de covariance et, par exemple, de calculer la moyenne d'un ensemble de matrices de covariance. Cela permet notamment de calculer la matrice de covariance moyenne pour chaque classe de signaux EEG (par exemple des mouvements imaginés de la main gauche ou des mouvements imaginés de la main droite), à partir de données d'apprentissage. Un classifieur peut ensuite être construit en utilisant ces matrices de covariance moyennes comme prototypes pour chaque classe. Ainsi, lorsque l'on veut classifier de nouveaux signaux EEG avec un MDRM, on filtre d'abord ces signaux EEG puis on calcule leur matrice de covariance. On attribue ensuite à cette matrice de covariance la classe dont la matrice de covariance moyenne est la plus proche, selon une distance spécifique : la distance riemannienne (cf figure 2). Ce classifieur est très simple, mais déjà très efficace, car 1) il ne requiert que peu de paramètres, et peut donc être appris à partir de peu de données d'apprentissage (et les expériences BCI ne permettent généralement de récolter qu'un petit nombre de données) et 2) la distance riemannienne possède des propriétés d'invariance intéressantes (elle est invariante par transformation linéaire de rang plein), qui rendent les BCI qui utilisent les CR plus robustes à certaines des variabilités affectant les signaux EEG.

Ces dernières années, de tels CR, et leurs variantes, se sont révélés particulièrement efficaces. Ce fut par exemple le cas lors d'une compétition internationale de classification de signaux EEG que nous avons organisée en 2021 [40]. Lors de cette compétition, les participants devaient estimer le niveau de charge mentale de personnes (c'est-à-dire les efforts mentaux qu'ils fournissent, parmi trois niveaux ou classes: charge mentale basse, moyenne ou élevée) à partir de leurs signaux EEG. Pour cela les participants avaient à leur disposition deux jours de signaux EEG pour chaque participant, avec les niveaux de charge mentale

correspondants. Ils devaient se servir de ces données comme ensemble d'apprentissage, pour entraîner et optimiser leurs classifieurs et leurs algorithmes d'apprentissage artificiel, qu'ils devaient ensuite utiliser pour estimer la charge mentale des mêmes personnes mais pour un jour différent. C'est un problème difficile car les signaux EEG sont connus pour changer beaucoup d'un jour à l'autre. Lors de cette compétition, les pires performances de classification obtenues (bonne reconnaissance des niveaux de charge mentale), qui étaient au niveau de la chance (c'est-à-dire que le classifieur était incapable d'estimer la charge mentale mieux qu'en devinant au hasard), ont toutes été obtenues, de manière surprenante, par des algorithmes d'apprentissage profond (*deep learning*). Contrairement à d'autres domaines tels que l'image ou la parole, l'apprentissage profond n'a pas encore révolutionné les BCI, probablement parce que ces algorithmes ont besoin d'un très grand nombre de données d'apprentissage (typiquement des millions), alors que les bases de données BCI ont seulement des centaines ou milliers de données en moyenne. Cela pourrait cependant changer dans le futur. Pour cette compétition, le gagnant a utilisé des CR, et parmi les 4 meilleures performances, 3 ont utilisé des CR [40]. Plus généralement, ces dernières années, parmi les 9 compétitions internationales de classification de signaux EEG que nous connaissons, 8 ont été gagnées par des CR, et une seule par de l'apprentissage profond. Ainsi, actuellement les classifieurs riemanniens sont considérés comme l'état de l'art pour la classification de signaux EEG en BCI.

Les CR étant l'état de l'art pour les BCI, nous les avons récemment beaucoup étudiés, et avons proposé de nombreuses variantes et améliorations pour ceux-ci. Par exemple, nous avons proposé des algorithmes pour sélectionner automatiquement les meilleures électrodes pour les CR [41], ainsi que des CR adaptatifs, capables de mettre à jour leurs paramètres de manière dynamique et automatique quand de nouvelles données sont disponibles [5, 23], des classifieurs multimodaux plus robustes aux bruits et variabilités [51], de l'apprentissage par transfert riemannien [9], une identification automatique des meilleures bandes de fréquence EEG pour les CR [3, 49], ou encore l'utilisation de mesures d'interactions entre bandes de fréquences EEG dans les CR [50]. Ainsi, il existe de nombreuses approches d'apprentissage artificiel qui permettent d'améliorer la fiabilité et la stabilité des BCI. Cependant, si celles-ci sont nécessaires, elles ne sont pas nécessairement suffisantes pour obtenir des BCI complètement robustes.

Apprentissage humain dans les BCI

Un autre élément clé de la boucle d'interaction cerveau-ordinateur est l'utilisateur lui-même. En effet, contrôler une BCI est une compétence qui s'apprend : plus on pratique le contrôle d'une BCI, meilleur on devient, et plus précisément nos commandes mentales sont reconnues par la BCI [39]. Ainsi, pour obtenir

une BCI robuste, il ne faut pas seulement de bons algorithmes de classification, il faut aussi un utilisateur qui soit bien entraîné. Pour cela il faut s'intéresser au feedback et aux tâches d'entraînement qui sont fournies aux utilisateurs de BCI, pour s'assurer d'un entraînement efficace et efficient.

L'entraînement classique au contrôle de BCI basées sur l'imagination de mouvements, qui est utilisé tel quel depuis une vingtaine d'années, consiste tout d'abord à indiquer à l'utilisateur à quel moment imaginer un mouvement, à l'aide d'une flèche affichée à l'écran (par exemple, une flèche à gauche ou à droite pour indiquer d'imaginer un mouvement de la main gauche ou droite, respectivement). Ensuite, lorsque l'utilisateur imagine ce mouvement, il reçoit un feedback sous forme d'une barre, grandissant vers la gauche ou vers la droite, dont la direction indique le mouvement qui est reconnu (gauche ou droite pour un mouvement imaginé de la main gauche ou droite, respectivement), et la taille indique la confiance du classifieur dans cette reconnaissance (cf. figure 1, par exemple). Cet entraînement est répété tel quel plusieurs fois par jour, pendant plusieurs jours, semaines voir mois. Si on étudie cette forme d'entraînement de manière plus formelle, on peut cependant se rendre compte qu'elle est sous-optimale, à la fois théoriquement mais aussi en pratique. En effet, d'un point de vue théorique, ce type d'entraînement ne respecte pas les principes et recommandations venant de la psychologie de l'apprentissage humain [26]. Par exemple, cet entraînement n'est pas adapté aux compétences et au profil de chaque utilisateur : il est le même pour tous, toujours. Cet entraînement est également sous-optimal en pratique. En effet, si on fournit à des utilisateurs le même exercice et le même feedback, mais pour apprendre autre chose que le contrôle de BCI, par exemple des tâches motrices simples, de nombreux utilisateurs vont également échouer à apprendre, car le feedback n'est pas assez informatif [20]. Ainsi, il est nécessaire de mieux comprendre, de modéliser et d'optimiser cet entraînement utilisateur en BCI.

Pour ce faire, à l'aide de nombreuses expérimentations couplées à une étude de la littérature, nous nous sommes intéressés à comprendre qui pouvait utiliser une BCI. En particulier, y a-t-il certains profils d'utilisateurs, en termes de personnalité ou de capacités cognitives, qui arrivent mieux que d'autres à contrôler une BCI ? Nos études nous ont permis d'identifier trois grandes familles de facteurs qui peuvent expliquer des différences entre utilisateurs [22, 21]. La première famille est formée des habiletés spatiales : capacité à créer, manipuler et transformer des images mentales, ce qui peut se mesurer avec un test de rotation mentale. Les utilisateurs avec de meilleures habiletés spatiales ont généralement de meilleures performances de contrôle des BCI. De manière similaire, les utilisateurs avec de bonnes capacités attentionnelles ou qui portent plus attention aux tâches d'entraînement, ont, eux aussi, de meilleures performances de contrôle des BCI. A l'inverse, les utilisateurs qui ont des problèmes de relations avec les technologies, par exemple qui souffrent d'anxiété

informatique ou qui ne se sentent pas en contrôle avec des technologies, ont généralement de moins bonnes performances de contrôle des BCI. Cela montre qu'il est nécessaire de prendre en compte le profil de l'utilisateur dans la conception de BCI et dans l'entraînement utilisateur.

En plus d'essayer de comprendre l'apprentissage humain en BCI, nous avons aussi cherché à l'optimiser, pour que les utilisateurs de BCI puissent apprendre à les contrôler plus rapidement ou plus efficacement. Nous avons ainsi exploré différents types de feedbacks et de tâches d'entraînement, par exemple [31, 37]. Pour ne détailler qu'un seul exemple, nous avons notamment proposé PEANUT (*Personalized emotional agent for neurotechnology user training*), un compagnon d'apprentissage artificiel pour guider et aider les utilisateurs dans leur apprentissage du contrôle d'une BCI [36] (cf. figure 3).

Fig. 3. Notre compagnon d'apprentissage au contrôle de BCI, PEANUT (le petit humanoïde en bas à gauche sur l'image), lors d'un entraînement au contrôle de BCI [36] (© Inria / Photo C. Morel).

Ce compagnon a accès à des mesures de performance et de progression de l'utilisateur, et en fonction de celles-ci, va pouvoir interagir avec cet utilisateur durant son entraînement pour le guider. Il va par exemple pouvoir féliciter l'utilisateur quand il progresse, l'encourager à persévérer ou encore lui suggérer de changer de stratégie mentale quand ses performances et sa progression ne sont pas bonnes. Nous avons comparé un entraînement BCI avec et sans PEANUT, et les résultats se sont avérés particulièrement intéressants. En effet, PEANUT permet bien d'améliorer les performances de contrôle d'une BCI, mais pas pour

tout le monde. En particulier, PEANUT améliore les performances de contrôle des utilisateurs qui ont une personnalité dite *non-autonome* en psychologie, c'est-à-dire qui préfèrent travailler en groupe. Par contre, PEANUT diminue les performances de ceux qui ont une personnalité *autonome*, qui préfèrent travailler seuls. Pour ces derniers, PEANUT n'aide pas, mais est plutôt un compagnon qui les dérange et les ennuie. Une de fois de plus, cela montre bien que l'entraînement au contrôle des BCI doit être personnalisé, et prendre en compte le profil des utilisateurs, c'est-à-dire leurs personnalités et leurs capacités cognitives.

Applications des BCI

Au delà de recherches plus fondamentales visant à améliorer la robustesse et l'utilisabilité des BCI, que ce soit en s'intéressant aux aspects machine ou aux aspects humains des BCI, nous nous intéressons aussi à leurs applications concrètes. Dans la suite de cette section, nous allons illustrer deux familles d'applications que nous explorons : 1) en santé numérique et 2) en neuroergonomie, c'est-à-dire dans l'étude du cerveau hors du laboratoire, dans des tâches de la vie quotidienne, par exemple pour l'étude de l'expérience utilisateur dans les signaux EEG [4].

BCI pour la santé numérique

Historiquement, les BCI ont vu le jour principalement pour devenir une technologie d'assistance ou de rééducation pour les personnes gravement paralysées [30, 35]. Nos travaux s'intéressent ainsi naturellement à ce domaine d'application.

Nous avons ainsi conçu une BCI pour un utilisateur tétraplégique — paralysé des 4 membres donc — que nous avons entraîné pendant plusieurs mois pour qu'il puisse contrôler un jeu vidéo de course de voiture entièrement grâce à ses signaux EEG avec une BCI [5]. Nous avons effectué ces travaux dans le cadre d'une compétition du Cybathlon, qui se veut être des jeux olympiques des technologies d'assistances, où des personnes en situation de handicap s'associent avec des concepteurs de technologies pour relever des défis d'assistance, tels que du contrôle d'exosquelette, de prothèses ou encore de BCI [47]. Même si nous n'avons pas gagné cette compétition, notre participation nous a permis de développer de nouveaux algorithmes de traitement des signaux EEG, ainsi que d'identifier un nouveau type d'apprentissage humain dans les BCI [5].

Nous nous sommes aussi intéressés à la rééducation motrice après un accident vasculaire cérébral (AVC) à l'aide de BCI. En effet, suite à un AVC, qui peut donner lieu à des lésions cérébrales dans les aires motrices du cerveau, des personnes peuvent se retrouver paralysées d'un membre. Et cette paralysie peut les empêcher de participer à de nombreux exercices de rééducation. Par contre,

même si ces patients post-AVC ne peuvent pas bouger physiquement, il est possible de détecter qu'ils veulent bouger, dans leurs signaux EEG, à l'aide d'une BCI. Quand une BCI détecte une telle intention de mouvement dans les signaux EEG, elle peut déclencher un feedback pour l'utilisateur, en lui montrant une main en train de bouger ou en bougeant physiquement la main de l'utilisateur à l'aide d'une orthèse robotisée, par exemple [35]. Des travaux et métaanalyses ont ainsi montré qu'un tel feedback, lié à l'intention de mouvement détectée dans les signaux EEG, permettait d'améliorer la récupération motrice, en aidant les participants à utiliser l'aire cérébrale touchée par l'AVC, stimulant ainsi la plasticité cérébrale dans cette aire [35, 10]. Nous cherchons donc aussi à contribuer à cette ligne de recherche, et avons proposé, en collaboration avec le CHU de Bordeaux, une BCI qui peut détecter des intentions de mouvement de la main gauche ou droite, et, suite à une telle détection, montrer sur un écran une image 3D de la main correspondante en train de bouger, et fournir des vibrations tactiles sur la main réelle du patient [37] (cf. figure 4). Ce système a été validé sur des personnes en bonne santé (sans AVC) et va prochainement être testé dans un essai clinique avec des patients AVC, afin d'identifier les profils de patients qui peuvent bénéficier d'une rééducation post-AVC à l'aide de BCI [43].

BCI pour la neuroergonomie

Au delà des applications médicales, les BCI peuvent aussi s'avérer très prometteuses pour le grand public, notamment en neuroergonomie. En particulier les BCI peuvent estimer, dans les signaux EEG de leurs utilisateurs, certains états mentaux de ceux-ci (par exemple un niveau attentionnel), de façon à pouvoir évaluer des systèmes interactifs ou adapter le contenu de l'application ou son interaction à l'état de l'utilisateur [4, 15]. On parle alors de BCI dites « passives » [53].

Ainsi nous avons développé des algorithmes et des BCI pour estimer le niveau de charge mentale des utilisateurs dans leurs signaux EEG [3, 32], et avons utilisé ces BCI pour évaluer où et quand des dispositifs ou tâches d'interaction 3D étaient cognitivement difficiles à utiliser [46, 17]. En collaboration avec l'ISAE-SupAéro, nous avons également pu montrer que l'on pouvait estimer la charge mentale de pilotes d'avion en plein vol, dans leurs signaux EEG [14]. Dans le même ordre d'idée, nous avons montré que l'on pouvait estimer (avec cependant des performances de reconnaissance modestes), dans les signaux EEG, des aspects du confort visuel lorsqu'un utilisateur regarde des images stéréoscopiques (i.e., des images perçues comme étant en relief) [16]. Tout cela fournit ainsi de nouveaux outils, complémentaires à ceux existants (par exemple, des questionnaires ou interviews des utilisateurs), pour évaluer l'expérience utilisateur (UX) avec des dispositifs d'interaction humain-machine [18].

Plus récemment, et au delà de l'interaction humain-machine, nous avons exploré si l'on pouvait estimer d'autres types d'états mentaux dans les signaux

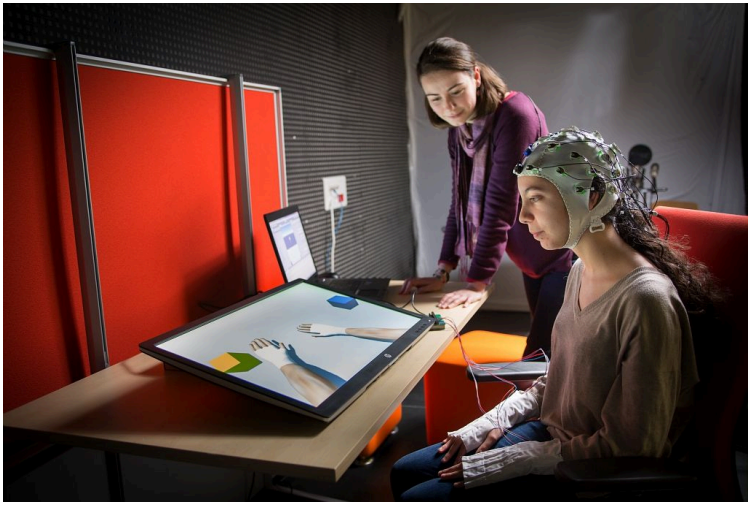


Fig. 4. Notre BCI pour la rééducation motrice post-AVC, fournissant un feedback visuel réaliste (mouvement d'une main en 3D) et un feedback vibrotactile sur la main, suite à la détection d'une intention de mouvement dans les signaux EEG [37] (© Inria / Photo C. Morel).

EEG. Nous avons ainsi étudié la curiosité, que l'on peut estimer — avec des performances modestes (on peut distinguer si un utilisateur est curieux ou pas avec une précision de l'ordre de 65 %) — dans les signaux EEG [2], ou encore l'expérience esthétique dans des musées virtuels (i.e. un utilisateur apprécie-t-il une œuvre d'art ou pas) [44, 45]. La possibilité de décoder de tels états mentaux dans les signaux EEG ouvre des perspectives intéressantes pour le futur, par exemple pour personnaliser des musées virtuels automatiquement pour chaque utilisateur (chaque utilisateur verra des œuvres d'arts différentes, qui maximiseront son expérience esthétique) ou pour personnaliser des dispositifs éducatifs pour chaque apprenant (chaque apprenant aura des séquences d'exercices différentes, qui maximiseront son engagement et sa curiosité pour améliorer l'apprentissage et la mémorisation.

Conclusion

Globalement, cet article a proposé un bref panorama de la recherche sur les interfaces cerveau-ordinateur (BCI) en général, et de nos travaux sur ce sujet en particulier. Il a notamment mis en avant le fait que les BCI sont très prometteuses pour de nombreuses applications, mais encore trop peu fiables pour être utilisées largement en pratique. Cet article a ainsi illustré nos travaux pour rendre ces BCI plus fiables, que ce soit en travaillant sur les aspects machine (algorithmes de

classification des signaux EEG), ou les aspects humains (comprendre et optimiser l'entraînement utilisateur). Il a enfin présenté quelques-uns de nos travaux applicatifs en BCI, dans le domaine de la santé numérique (pour les technologies d'assistance ou la rééducation motrice post-AVC) ou de la neuroergonomie (pour évaluer des états mentaux liés à l'expérience utilisateur, tels que la charge mentale ou la curiosité). Bien entendu, ce panorama est loin d'être complet, et n'est qu'un aperçu. Le lecteur intéressé pourra se référer notamment à [33, 11, 12] pour plus d'informations sur le sujet. Pour ceux qui souhaiteraient aller encore plus loin, et essayer les BCI ou les utiliser dans leurs travaux et recherches, nous pouvons leur recommander d'essayer le logiciel OpenViBE, un logiciel libre et gratuit que nous codéveloppons et qui permet de concevoir et d'utiliser des BCI en temps réel, et ce, sans programmer une seule ligne de code [38].

Malgré l'accélération des travaux sur les BCI, et tous les résultats de recherche récents sur le sujet, les BCI actuelles restent cependant encore insuffisamment fiables, notamment pour une utilisation hors des laboratoires. Il y a donc toujours besoin de recherches sur tous les composants des BCI, pour les rendre plus robustes, fiables et utilisables. Une direction de recherche potentiellement intéressante sur le sujet serait de développer des théories de l'interaction cerveau-ordinateur. En effet, la recherche actuelle en BCI est une recherche très expérimentale, procédant le plus souvent par essais et erreurs : un nouvel algorithme ou un nouveau protocole d'entraînement est proposé et testé, et on observe expérimentalement s'il donne satisfaction ou pas. Cette approche est déjà très utile, mais n'est pas très efficace. De plus, qu'une nouvelle méthode fonctionne ou pas, très souvent on ne sait pas pourquoi. Cela empêche d'accumuler les résultats et les savoirs et de pouvoir proposer des solutions optimales. Il y a donc un besoin de théories, pouvant expliquer et prédire le comportement des BCI et de ses composantes, par exemple, d'expliquer et prédire pourquoi et quand un algorithme de classification va pouvoir classifier les signaux EEG de manière fiable. De telles théories permettraient ensuite de concevoir de nouvelles approches et de nouveaux algorithmes qui soient optimaux par rapport à ces théories, et donc des BCI qui soient enfin fiables. C'est ce que nous nous attellerons à faire dans les années à venir.

Remerciements

Le travail présenté dans cet article a été financé en partie par les projets ANR REBEL (ANR-15-CE23-0013-01), ANR PROTEUS (ANR-22-CE33-0015-01), ERC BrainConquest (ERC-2016-STG-714567) et CHIST-ERA BITSCOPE (ANR-21-CHRA-0003-01). Je souhaite aussi remercier très chaudement tous mes nombreux collaborateurs et collaboratrices (trop nombreux pour être tous cités ici) qui ont permis à ces travaux de voir le jour.

Références

[1] Brendan Allison and Christa Neuper. 2010. Could anyone use a BCI? In *Brain-Computer Interfaces: Human-Computer Interaction Series*. 35–54. https://doi.org/10.1007/978-1-84996-272-8_3.

[2] Aurélien Appriou, Jessy Ceha, Smeety Pramij, Dan Dutartre, Edith Law, Pierre-Yves Oudeyer, and Fabien Lotte. 2020. Towards measuring states of epistemic curiosity through electroencephalographic signals. In *2020 ieee international conference on systems, man, and cybernetics (smc)*, 4006–4011.

[3] Aurélien Appriou, Andrzej Cichocki, and Fabien Lotte. 2020. Modern machine-learning algorithms for classifying cognitive and affective states from electroencephalography signals. *IEEE Systems, Man, and Cybernetics Magazine* 6, 3:29–38.

[4] Hasan Ayaz and Frédéric Dehais. 2018. *Neuroergonomics: The brain at work and in everyday life*. Academic Press.

[5] Camille Benaroch, Khadijeh Sadatnejad, Aline Roc, Aurélien Appriou, Thibaut Monseigne, Smeety Pramij, Jelena Mladenovic, Léa Pillette, Camille Jeunet, and Fabien Lotte. 2021. Long-Term BCI Training of a Tetraplegic User: Adaptive Riemannian Classifiers and User Training. *Frontiers in Human Neuroscience* 15. https://doi.org/10.3389/fnhum.2021.635653.

[6] N Birbaumer, N Ghanayim, Thilo Hinterberger, Iver Iversen, Boris Kotchoubey, Andrea Kübler, J Perelmouter, E Taub, and H Flor. 1999. A spelling device for the paralysed. *Nature* 398:297–8. https://doi.org/10.1038/18581.

[7] Benjamin Blankertz, Claudia Sannelli, Sebastian Halder, Eva M Hammer, Andrea Kübler, Klaus-Robert Müller, Gabriel Curio, and Thorsten Dickhaus. 2010. Neurophysiological predictor of SMR-based BCI performance. *Neuroimage* 51, 4:1303–1309.

[8] Laurent Bonnet, Fabien Lotte, and Anatole Lécuyer. 2013. Two brains, one game: Design and evaluation of a multiuser BCI video game based on motor imagery. *IEEE Transactions on Computational Intelligence and AI in games* 5, 2:185–198.

[9] Laurent Bougrain, Sébastien Rimbert, Pedro LC Rodrigues, Geoffrey Canron, and Fabien Lotte. 2021. Guidelines to use transfer learning for motor imagery detection: An experimental study. In *2021 10th international IEEE/EMBS conference on neural engineering (NER)*, 5–8.

[10] María A Cervera, Surjo R Soekadar, Junichi Ushiba, José del R Millán, Meigen Liu, Niels Birbaumer, and Gangadhar Garipelli. 2018. Brain-computer interfaces for post-stroke motor rehabilitation: A meta-analysis. *Annals of clinical and translational neurology* 5, 5:651–663.

[11] Maureen Clerc, Laurent Bougrain, and Fabien Lotte. 2016. *Les interfaces cerveau-ordinateur 1 : Fondements et méthodes*. ISTE Group.

[12] Maureen Clerc, Laurent Bougrain, and Fabien Lotte. 2016. *Les interfaces cerveau-ordinateur 2 : Technologie et applications*. ISTE Group.

[13] Marco Congedo, Alexandre Barachant, and Rajendra Bhatia. 2017. Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review. *Brain-Computer Interfaces* 4, 3:155.

[14] Frédéric Dehais, Alban Duprès, Sarah Blum, Nicolas Drougard, Sébastien Scannella, Raphaëlle N Roy, and Fabien Lotte. 2019. Monitoring pilot's mental workload using ERPs and spectral power with a six-dry-electrode EEG system in real flight conditions. *Sensors* 19, 6:1324.

[15] Stephen H Fairclough and Fabien Lotte. 2020. Grand challenges in neurotechnology and system neuroergonomics. *Frontiers in Neuroergonomics 1*, 602504.

[16] Jérémy Frey, Aurélien Appriou, Fabien Lotte, and Martin Hachet. 2016. Classifying EEG signals during stereoscopic visualization to estimate visual comfort. *Computational intelligence and neuroscience* 2016, 1:2758103.

[17] Jérémy Frey, Maxime Daniel, Julien Castet, Martin Hachet, and Fabien Lotte. 2016. Framework for electroencephalography-based evaluation of user experience. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, 2283–2294.

[18] J Frey 1 2, M Hachet 2, and F Lotte 2. 2017. EEG-based neuroergonomics for 3D user interfaces: Opportunities and challenges. *Le travail humain* 80, 1:73–92.

[19] Rümeysa İnce, Saliha Seda Adanır, and Fatma Sevmez. 2021. The inventor of electroencephalography (EEG): Hans berger (1873–1941). *Child's Nervous System* 37:2723–2724.

[20] Camille Jeunet, Emilie Jahanpour, and Fabien Lotte. 2016. Why standard brain-computer interface (BCI) training protocols should be changed: An experimental study. *Journal of Neural Engineering* 13. https://doi.org/10.1088/1741-2560/13/3/036024.

[21] Camille Jeunet, Fabien Lotte, Jean-Marie Batail, Pierre Philip, and Jean-Arthur Micoulaud Franchi. 2018. Using recent BCI literature to deepen our understanding of clinical neurofeedback: A short review. *Neuroscience* 378:225–233.

[22] C. Jeunet, B. N'Kaoua, and F. Lotte. 2016. Advances in user-training for mental-imagery-based BCI control: Psychological and cognitive factors and their neural correlates. In *Brain-computer interfaces: Lab experiments to real-world applications*, Damien Coyle (ed.). Elsevier, 3–35. https://doi.org/https://doi.org/10.1016/bs.pbr.2016.04.002.

[23] Satyam Kumar, Florian Yger, and Fabien Lotte. 2019. Towards adaptive classification using riemannian geometry approaches in brain-computer interfaces. In *2019 7th international winter conference on brain-computer interface (BCI)*, 1–6.

[24] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. 2018. A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update. *Journal of neural engineering* 15, 3:031005.

[25] Fabien Lotte, Marco Congedo, Anatole Lécuyer, and Lamarche Fabrice. 2007. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering* 4.

[26] Fabien Lotte, Larrue Florian, and Christian Mühl. 2013. Flaws in current human training protocols for spontaneous brain-computer interfaces: Lessons learned from instructional design. *Frontiers in human neuroscience* 7:568. https://doi.org/10.3389/fnhum.2013.00568.

[27] Fabien Lotte and Cuntai Guan. 2011. Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms. *IEEE Transactions on biomedical Engineering* 58, 2:355–362.

[28] Fabien Lotte, Anatole Lécuyer, Fabrice Lamarche, and Bruno Arnaldi. 2007. Studying the use of fuzzy inference systems for motor imagery classification. *IEEE transactions on neural systems and rehabilitation engineering* 15, 2:322–324.

[29] Fabien Lotte, Aurelien Van Langhenhove, Fabrice Lamarche, Thomas Ernest, Yann Renard, Bruno Arnaldi, and Anatole Lécuyer. 2010. Exploring large virtual environments by thoughts using a brain–computer interface based on motor imagery and high-level commands. *Presence: teleoperators and virtual environments* 19, 1:54–70.

[30] J d R Millán, Rüdiger Rupp, Gernot R Müller-Putz, Roderick Murray-Smith, Claudio Giugliemma, Michael Tangermann, Carmen Vidaurre, Febo Cincotti, Andrea Kübler, Robert Leeb, and others. 2010. Combining brain–computer interfaces and assistive technologies: State-of-the-art and challenges. *Frontiers in neuroscience* 4:161.

[31] Jelena Mladenović, Jérémy Frey, Smeety Pramij, Jérémie Mattout, and Fabien Lotte. 2022. Towards Identifying Optimal Biased Feedback for Various User States and Traits in Motor Imagery BCI. *IEEE Transactions on Biomedical Engineering* 69, 3:1101–1110.

[32] Christian Mühl, Camille Jeunet, and Fabien Lotte. 2014. EEG-based workload estimation across affective contexts. *Frontiers in neuroscience* 8:114.

[33] Chang S Nam, Anton Nijholt, and Fabien Lotte. 2018. *Brain–computer interfaces handbook: Technological and theoretical advances*. CRC Press.

[34] Ernst Niedermeyer. 2011. *Niedermeyer's electroencephalography: Basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.

[35] Floriana Pichiorri, Giovanni Morone, Manuela Petti, Jlenia Toppi, Iolanda Pisotta, Marco Molinari, Stefano Paolucci, Maurizio Inghilleri, Laura Astolfi, Febo Cincotti, and others. 2015. Brain–computer interface boosts motor imagery practice during stroke recovery. *Annals of neurology* 77, 5:851–865.

[36] Léa Pillette, Camille Jeunet, Boris Mansencal, Roger N'Kambou, Bernard N'Kaoua, and Fabien Lotte. 2020. A physical learning companion for mental-imagery BCI user training. *Int. J. Hum.-Comput. Stud.* 136, C. https://doi.org/10.1016/j.ijhcs.2019.102380.

[37] Léa Pillette, Bernard N'kaoua, Romain Sabau, Bertrand Glize, and Fabien Lotte. 2021. Multi-session influence of two modalities of feedback and their order of presentation on MI-BCI user training. *Multimodal Technologies and Interaction* 5, 3:12.

[38] Yann Renard, Fabien Lotte, Guillaume Gibert, Marco Congedo, Emmanuel Maby, Vincent Delannoy, Olivier Bertrand, and Anatole Lécuyer. 2010. OpenViBE: An open-source software platform to design, test, and use brain–computer interfaces in real and virtual environments. *Presence* 19, 1:35–53.

[39] Aline Roc, Léa Pillette, Jelena Mladenović, Camille Benaroch, Bernard N'Kaoua, Camille Jeunet, and Fabien Lotte. 2021. A review of user training methods in brain computer interfaces based on mental tasks. *Journal of Neural Engineering* 18, 1:011002. https://doi.org/10.1088/1741-2552/abca17.

[40] Raphaëlle N. Roy, Marcel F. Hinss, Ludovic Darmet, Simon Ladouce, Emilie S. Jahanpour, Bertille Somon, Xiaoqi Xu, Nicolas Drougard, Frederic Dehais, and Fabien Lotte. 2022. Retrospective on the First Passive Brain-Computer Interface Competition on Cross-Session Workload Estimation. *Frontiers in Neuroergonomics* 3. https://doi.org/10.3389/fnrgo.2022.838342.

[41] Khadijeh Sadatnejad and Fabien Lotte. 2022. Riemannian channel selection for BCI with between-session non-stationarity reduction capabilities. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 30:1158–1171.

[42] Simanto Saha and Mathias Baumert. 2020. Intra- and inter-subject variability in EEG-based sensorimotor brain computer interface: A review. *Frontiers in Computational Neuroscience* 13.

[43] David Trocellier, Bernard N'kaoua, and Fabien Lotte. 2022. Identifying factors influencing the outcome of BCI-based post stroke motor rehabilitation towards its personalization with artificial intelligence. In *2022 IEEE international conference on metrology for extended reality, artificial intelligence and neural engineering (MetroXRAINE)*, 151–156.

[44] Marc Welter and Fabien Lotte. 2024. Ecological decoding of visual aesthetic preference with oscillatory electroencephalogram features—a mini-review. *Frontiers in Neuroergonomics* 5. https://doi.org/10.3389/fnrgo.2024.1341790.

[45] Marc Welter, J Casal Martínez, E Redmond, Jonathan Baum, Tomas Ward, and Fabien Lotte. 2024. EEG-based art interest decoding. In *Neuroergonomics conference 2024*.

[46] Dennis Wobrock, Jérémy Frey, Delphine Graeff, Jean-Baptiste de La Rivière, Julien Castet, and Fabien Lotte. 2015. Continuous mental effort evaluation during 3D object manipulation tasks based on brain and physiological signals. In *Human-computer interaction–INTERACT 2015: 15th IFIP TC 13 international conference, bamberg, germany, september 14-18, 2015, proceedings, part i 15*, 472–487.

[47] Peter Wolf and Robert Riener. 2018. Cybathlon: How to promote the development of assistive technologies. *Science robotics* 3(17):eaat7174.

[48] Jonathan R Wolpaw and E Winter Wolpaw. 2012. Brain-computer interfaces: Something new under the sun. *Brain-computer interfaces: principles and practice* 14:3–12.

[49] Maria Sayu Yamamoto, Fabien Lotte, Florian Yger, and Sylvain Chevallier. 2022. Class-distinctiveness-based frequency band selection on the riemannian manifold for oscillatory activity-based BCIs: Preliminary results. In *2022 44th annual international conference of the IEEE engineering in medicine & biology society (EMBC)*, 3690–3693.

[50] Maria Sayu Yamamoto, Apolline Mellot, Sylvain Chevallier, and Fabien Lotte. 2023. Novel SPD matrix representations considering cross-frequency coupling for EEG classification using riemannian geometry. In *2023 31st european signal processing conference (EUSIPCO)*, 960–964.

[51] Maria Sayu Yamamoto, Khadijeh Sadatnejad, Toshihisa Tanaka, Md Rabiul Islam, Frédéric Dehais, Yuichi Tanaka, and Fabien Lotte. 2023. Modeling complex EEG data distribution on the riemannian manifold toward outlier detection and multimodal classification. *IEEE Transactions on Biomedical Engineering*.

[52] Florian Yger, Maxime Berar, and Fabien Lotte. 2017. Riemannian approaches in brain-computer interfaces: A review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25, 10:1753–1762.

[53] Thorsten O Zander and Christian Kothe. 2011. Towards passive brain–computer interfaces: Applying brain–computer interface technology to human–machine systems in general. *Journal of neural engineering* 8, 2:025005.

[54] Mingjun Zhong, Fabien Lotte, Mark Girolami, and Anatole Lécuyer. 2008. Classifying EEG for brain computer interfaces using gaussian processes. *Pattern Recognition Letters* 29, 3:354–359.

Algorithmes d'approximation et sketches pour les problèmes de clustering

David Saulpic[1]

Cet article est le résumé de la thèse de doctorat[2] *de son auteur, qui a reçu le prix de thèse Gilles Kahn 2023, décerné par la Société informatique de France.*

Le titre de ma thèse cache plusieurs termes techniques, que je vais m'attacher à expliquer un à un, pour motiver les questions abordées. Mais tout d'abord, un peu de contexte. Ces travaux s'inscrivent dans le domaine de l'informatique théorique ; les problèmes que nous avons étudiés ont leurs racines dans le traitement des « big data », mais notre apport est avant tout théorique : avec quelle précision peut-on espérer répondre à ces problèmes, même dans des cas très contraints ? Comment simplifier leur structure et les données d'entrée avant de répondre au problème ? L'objectif premier n'est pas d'implémenter, mais d'abord de comprendre — ce qui, parfois, mène à des implémentations efficaces, comme nous avons pu l'expérimenter dans ce travail.

Les problèmes de clustering

Ce nom peut avoir de nombreuses interprétations, et cache différents problèmes. Pendant ma thèse, nous nous sommes concentrés principalement sur les problèmes des k-moyennes et k-médianes. Leur but est de répondre à la question

1. Institut de recherche en informatique fondamentale (IRIF), université Paris Cité & CNRS.
2. https://theses.fr/2022SORUS194.

DOI:10.48556/SIF.1024.24.145

suivante : comment compresser un nuage de points de façon fidèle ? Si l'objectif est de représenter le nuage par un point unique, la solution habituelle est de choisir la moyenne — voir Figure 1. Si les points sont simplement des nombres réels (i.e., le nuage de points est de dimension 1), alors on peut aussi choisir la médiane.

Mais si on s'autorise à représenter le nuage par plus d'un point, disons k points, comment les choisir ? Pour formaliser le problème, étant donné un ensemble de points $P \subset \mathbb{R}^d$, on cherche à calculer un ensemble S de k représentants qui soit *fidèle*. Au début des années 60, une fonction de coût a été définie (notamment par Lloyd [4]) de la façon suivante : si on considère que chaque point de P est remplacé par son représentant le plus proche, l'erreur faite est la distance entre le point et son représentant. Le *coût* d'un ensemble de représentants S est alors la somme des erreurs au carré. En équation,

$$\mathrm{cost}(P, S) = \sum_{p \in P} \min_{s \in S} \mathrm{dist}(p, s)^2. \tag{1}$$

Le but du problème des k-moyennes est de calculer l'ensemble S de taille k qui minimise la fonction cost, avec l'espoir qu'un faible coût implique une bonne représentativité. Le nom vient du fait que le problème généralise la moyenne : pour $k = 1$, le point minimisant le coût est la moyenne de P. Le problème des k-médianes est similaire, mais le coût est simplement la somme des erreurs. Cette fonction objectif généralise la médiane : si $k = 1$ et $P \subset \mathbb{R}$, alors la meilleure solution est la médiane.

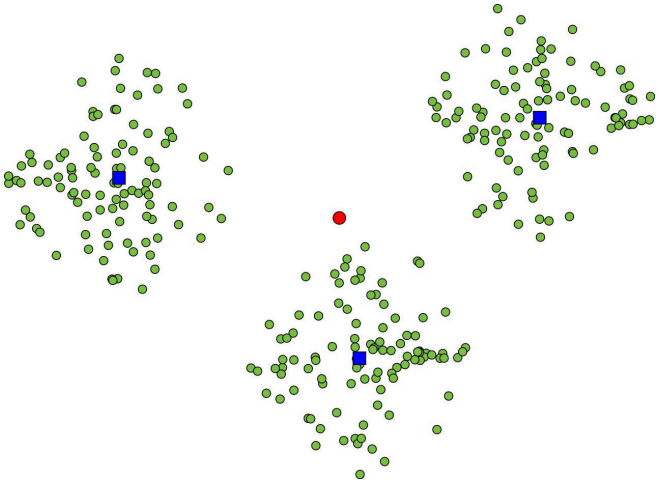


Fig. 1. Un nuage de points avec trois groupes distincts. La moyenne, le cercle rouge, est un point artificiel ne correspondant à rien dans le nuage de point. En revanche, les trois carrés bleus représentent plus correctement le nuage.

L'intérêt de compresser un nuage de points par k points plutôt qu'un seul est évident dès que les points sont regroupés en *clusters*, comme dans l'exemple présenté en figure 1 : dans ce cas, il y a trois groupes de points distincts, et choisir $k = 3$ permet une bien meilleure représentation que de garder simplement la moyenne. En un sens, les représentants permettent de détecter les clusters du nuage, d'où le nom *clustering*.

Algorithmes d'approximation

Malheureusement, le problème de minimiser la fonction cost est NP-complet [5]. À la place, on peut chercher à calculer une solution aussi bonne que possible : plutôt que de chercher la solution optimale, on peut essayer de calculer une solution qui a un coût $(1+\varepsilon)$ fois l'optimal. On appelle une telle solution une $(1+\varepsilon)$-approximation. Plus ε est petit, meilleure est la solution : la grande question des algorithmes d'approximation est de savoir à quel point ε peut être proche de zéro. Cette étude du problème permet de mieux comprendre sa structure, et de développer des outils techniques qui sont souvent utilisables dans d'autres scénarios.

Pour les k-moyennes, un résultat récent [1] montre qu'un algorithme simple dit de *recherche locale* permet de calculer une $(1+\varepsilon)$-approximation pour n'importe quel ε fixé, au prix d'un temps de calcul $nk(\log n)^{(d/\varepsilon)^{O(d)}}$. Dans la première partie de ma thèse, nous avons eu deux objectifs : d'abord, réduire ce temps de calcul le plus possible ; ensuite, étendre le résultat à des espaces métriques les plus généraux possibles.

En effet, P n'est pas forcément un nuage de points dans un espace euclidien (par exemple, la distance entre points n'est pas forcément la distance euclidienne mais peut être une autre norme ℓ_p). Nous avons cherché à réduire le plus possible les hypothèses sur P, pour identifier quelle propriété de P rend le problème « facile ». La propriété que nous avons identifiée est la suivante : il faut que, dans l'espace métrique, toute boule de rayon R soit couverte par 2^d boules de rayon $R/2$ — pour tout R. On dit alors que la *doubling dimension* de l'espace est d. On a construit un algorithme qui permet, dans de tels espaces, de calculer une $(1+\varepsilon)$ approximation avec un temps de calcul essentiellement $n \cdot 2^{1/\varepsilon^{d^2}}$. Quand n et k sont grands, c'est beaucoup mieux que le temps de calcul précédent — mais le résultat reste théorique et l'algorithme éloigné de toute application !

En revanche, les techniques et résultats théoriques que l'on a développés lors de ce travail trouvent des applications plus pratiques : on a montré comment les utiliser pour développer un algorithme efficace en pratique, et qui a en plus des garanties de *confidentialité différentielle* — d'une certaine manière, cet algorithme respecte la confidentialité des utilisateurs et utilisatrices qui ont fourni les données du nuage de points.

Sketches

Le dernier mot clef du titre concerne la seconde partie de ma thèse. En informatique théorique, un *sketch* (« esquisse » en français) est un résumé de l'entrée, qui utilise peu de mémoire mais permet de calculer une solution (approchée) au problème considéré. Le but ici est de construire des sketches qui utilisent le moins de mémoire possible, par exemple pour résoudre le problème dans des cas de « big data » où l'entrée du problème ne tient pas dans la mémoire d'une unique machine. En d'autres mots, pour le problème des k-moyennes, on cherche à compresser l'entrée pour pouvoir trouver une solution qui minimise le coût : donc, on compresse pour pouvoir compresser ! Mais ce procédé ne se mord pas la queue, parce que les propriétés désirables de la première compression sont bien différentes de celles de la seconde.

Une forme de sketches, celle que j'ai le plus étudié, s'appelle un *coreset*. On dit que Ω est un ε-coreset pour P si, pour tout ensemble S de k représentants,

$$\mathrm{cost}(P, S) = (1 \pm \varepsilon)\mathrm{cost}(\Omega, S). \tag{2}$$

Si Ω contient peu de points différents, on peut donc oublier P et résoudre le problème des k-moyennes directement sur Ω. Un des principaux résultats de ma thèse est de caractériser des propriétés de P qui permettent de calculer un petit coreset. Si $P \subset \mathbb{R}^d$, on a montré l'existance de coreset de taille $k\varepsilon^{-4}$ — donc indépendant de la taille de P ! Pour un k fixé et une précision ε donnée, n'importe quel ensemble de points P, qu'il contienne 10^6 ou 10^{100} points, peut se résumer en un ensemble de même taille. Cette propriété était déjà connue (voir par exemple [3]) : notre contribution est de réduire la taille du coreset, et ce dans un cadre très général : par exemple, notre résultat s'étend directement si l'entrée à une faible doubling dimension, ou bien si la distance entre les points est donnée par un graphe structuré.

Perspectives

Les contributions de ma thèse sont avant tout de nature théorique, mais les problèmes de k-moyennes et k-médianes font partie des principaux outils à la fois du traitement de « big data » (qu'ils permettent de compresser) et d'apprentissage non-supervisé, car ils permettent une classification de l'entrée. Pour de nombreuses raisons, je ne me reconnais pas forcément dans ces applications. Notamment, je me méfie d'un problème identifié par exemple par Cathy O'Neil [6] : les algorithmes traitant des données à grande échelle peuvent rendre dramatique le moindre biais, ou la moindre fuite de confidentialité.

J'essaie de réorienter ma recherche vers de nouvelles perspectives, en combinant l'informatique théorique avec d'autres disciplines. Par exemple, des contraintes légales ou sociétales imposent désormais aux algorithmes une

certaine confidentialité et une certaine équité. Je cherche actuellement à développer des algorithmes qui respectent ces contraintes, et aimeraisquestionner les techniques et définitions actuelles pour en comprendre la portée et les limites. J'aimerais aussi utiliser les outils et la manière de penser de l'informatique théorique dans d'autres disciplines. Par exemple, avec un groupe de chercheurs et chercheuses de Sorbonne Université, l'université de Paris-Cité et l'université Panthéon-Assas, nous menons des recherches sur le découpage de la France en circonscriptions pour les élections législatives, et essayons de mesurer la représentativité du découpage actuel [2].

Références

[1] Vincent Cohen-Addad. 2018. A fast approximation scheme for low-dimensional k-means. In (SODA '18). Retrieved from `http://dl.acm.org/citation.cfm?id=3174304.3175298`.

[2] Thomas Ehrhard, Solal Attias, Evripidis Bampis, Vincent Cohen-Addad, Bruno Escoffier, Claire Mathieu, Fanny Pascual, Adèle Pass-Lanneau, and David Saulpic. 2022. Découpage électoral des circonscriptions législatives en france. *Revue francaise de science politique* 72, 3:333–364.

[3] Dan Feldman and Michael Langberg. 2011. A unified framework for approximating and clustering data. In *Proceedings of the 43rd ACM symposium on theory of computing, STOC 2011, san jose, CA, USA, 6-8 june 2011*, 569–578.

[4] Stuart Lloyd. 1982. Least squares quantization in PCM. *IEEE transactions on information theory* 28, 2:129–137.

[5] Nimrod Megiddo and Kenneth J. Supowit. On the complexity of some common geometric location problems. *SIAM J. Comput.* 13.

[6] Cathy O'Neil. 2018. *Algorithmes, la bombe à retardement*. Les arènes.

La vie secrète des images JPEG

Détection de falsification via les traces de compression

Tina Nikoukhah

Cet article est le résumé de la thèse de doctorat[1] de son auteur, qui a reçu un accessit du prix de thèse Gilles Kahn 2023, décerné par la Société informatique de France.

Lors de la capture d'une image numérique, plusieurs étapes de traitement sont nécessaires avant d'obtenir le fichier final au format JPEG. Le capteur de l'appareil photo enregistre d'abord des données brutes, qui sont ensuite soumises à divers traitements, tels que le dématriçage, la balance des blancs et les corrections optiques. Enfin, l'image est compressée au format JPEG, un processus largement utilisé pour réduire la taille des fichiers, mais qui entraîne une perte d'information. Cette perte se manifeste souvent par l'apparition de blocs carrés, appelés artefacts de compression. Ces artefacts, caractéristiques de la compression JPEG, peuvent être analysés pour détecter des retouches ou falsifications, car toute modification locale de l'image perturbe l'homogénéité des traces de compression.

Contexte et motivation

Lors de la prise de vue avec un appareil photographique, les données brutes subissent généralement plusieurs traitements avant d'arriver à l'image finale. Une photographie numérique est débruitée, dématricée, corrigée des aberrations

1. https://theses.fr/2022UPASM027.

DOI:10.48556/SIF.1024.24.151

chromatiques et des distorsions optiques, puis subit des transformations non linéaires pour rehausser son contraste. Enfin, elle est compressée pour faciliter le stockage et la transmission. Le standard JPEG (*Joint Photographic Experts Group*) est le format de compression d'images numériques le plus couramment utilisé aujourd'hui pour les photographies. Cette compression entraîne une perte d'information et donc de qualité, se traduisant par l'apparition d'artefacts sous la forme de carrés, formant ce qu'on appelle la grille JPEG, comme illustré dans la figure 1.

Les étapes de compression JPEG sont décrites brièvement ci-après. Une transformation de l'espace de couleur est suivie d'un sous-échantillonnage des canaux de chrominance. Chaque canal est ensuite partitionné en blocs non chevauchants de 8×8 pixels et la transformée en cosinus discrète (DCT) de type II en 2D est appliquée à chacun de ces blocs. Chaque bloc de 8×8 pixels subit une étape de quantification effectuée dans le domaine spectral. Une table de quantification (liée à la qualité de compression, notée QF, allant de 1 à 100) fournit un facteur de quantification pour chaque composante DCT. Les artefacts JPEG sont clairement visibles lorsque l'image a été fortement compressée et sont presque imperceptibles lorsque la qualité de compression est élevée (i.e. le taux de compression est faible). Cependant, une grille est toujours présente en compression avec perte ($QF \leq 99$). À cette étape, certains coefficients DCT sont annulés lorsqu'ils ont une petite valeur par rapport au facteur de quantification. Ainsi, tous les blocs de taille 8×8 ont un nombre de zéros qui dépend à la fois de la qualité de compression et du contenu de l'image. Enfin, les coefficients DCT quantifiés sont compressés sans perte en exploitant, entre autres, la présence de valeurs nulles.

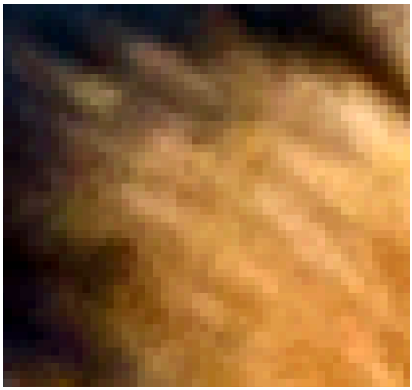


Fig. 1. Image non compressée.

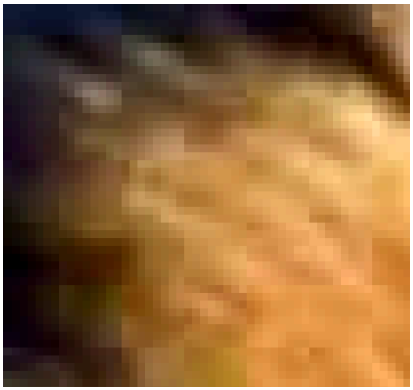


Fig. 2. Image fortement compressée.

À ces opérations classiques dans la chaîne de traitement d'une image peuvent se rajouter des modifications globales comme des changements de couleur, de luminosité ou des retouches locales, appelées falsifications. Ces opérations sont à la portée de tous grâce aux logiciels de retouche manuelle tels que *Gimp* ou *Photoshop* mais également ceux prenant en entrée du texte, tels que *Stable Diffusion* ou *Midjourney*. Le gommage, la suppression, le clonage ou l'insertion d'objets dans une image causent une rupture d'homogénéité des traces de compression. Pour détecter une falsification, les algorithmes basés sur la compression JPEG identifient des grilles locales non alignées avec la grille globale de l'image, ou des zones où la grille est absente. De plus, une zone retouchée peut être représentée comme une zone ayant une table de quantification différente de celle de l'en-tête du fichier ou de celle estimée globalement.

L'état de l'art de la détection de falsification s'appuie sur ces considérations pour proposer des révélateurs numériques [5], [8], [10], [11], [20], [3], [7], à savoir des opérateurs capables de mettre en évidence les zones semblant avoir été falsifiées. Un examen de ces méthodes montre néanmoins qu'elles souffrent de lacunes dans l'évaluation de la confiance qui peut être attribuée à leurs détections. D'où la nécessité de nouvelles méthodes donnant des résultats limitant le nombre de fausses alarmes.

Pour les algorithmes développés pendant ce travail de thèse, j'accompagne les détections dans l'image d'une validation statistique basée sur les méthodes *a contrario* [4]. Ce procédé permet de définir des détections à travers la probabilité que de tels événements se produisent par hasard, afin d'éviter les faux positifs. Ainsi, mes outils proposent une analyse automatique des images, ne nécessitant ni interprétation ni expertise dans le domaine.

Résultats principaux

L'historique de compression d'une image est définie comme l'origine de la grille globale et la table de quantification de toute compression subie par l'image. Si l'image est au format JPEG, ces informations doivent être comparées à celles de l'en-tête du fichier. Dans le cas où l'image ne serait pas au format JPEG, ce type d'analyse devient indispensable pour retrouver les informations d'une éventuelle compression JPEG antérieure.

Je propose une méthode, appelée *Qtable* [18], pour estimer la table de quantification JPEG qui se base uniquement sur les informations de l'image, sans utiliser de données provenant de l'en-tête du fichier. Pour cela, les coefficients DCT du canal de luminance de l'image sont analysés. La méthode se concentre sur les 63 coefficients AC et laisse le coefficient DC (qui a des propriétés différentes) non traité. L'histogramme de chacun des 63 coefficients est analysé et chaque valeur de quantification q entre 1 et 255 est évaluée. Ainsi, la méthode

proposée peut être appliquée à un format d'image non compressée pour détecter une précédente compression JPEG.

L'analyse des coefficients DCT et de leurs histogrammes révèle que de nombreuses valeurs ont été mises à zéro. En effet, c'est la principale raison pour laquelle la compression JPEG est efficace pour stocker une image tout en réduisant l'espace de stockage. La méthode proposée dans [16] repose sur cette observation et peut détecter la présence et les coordonnées des origines de grille globale d'une image. Par elle-même, cette méthode peut détecter si une image ayant subi une compression JPEG a été recadrée. De plus, elle peut être employée en tant qu'étape préliminaire à toute méthode de détection de falsification d'images, car nous avons observé que le fait de connaître l'origine de la grille de l'image et de la réaligner en conséquence améliore l'efficacité de ces méthodes.

Dans le cas d'une image falsifiée, les méthodes décrites précédemment indiquent que l'image a subi plusieurs compressions, car la table Q estimée est différente de celle qui est stockée dans l'en-tête du fichier et qu'il y a plusieurs origines de grille significatives. Cela n'atteste pas une falsification de l'image, mais plutôt fournit des informations intéressantes sur son historique, susceptibles de susciter des doutes. Pour détecter une altération, une approche consiste à repérer des grilles locales qui ne sont pas alignées avec la grille globale de l'image, ou à identifier une zone où la grille est absente. De plus, une zone retouchée peut se manifester par une table de quantification différente de celle mentionnée dans l'en-tête du fichier ou de celle estimée globalement.

La première approche est présentée dans [19] et [17]. La méthode exploite les traces d'artefacts en blocs pour récupérer localement la grille intégrée dans l'image par la compression JPEG. L'algorithme renvoie une liste de grilles associées à chaque partie de l'image. Le seul paramètre qui nécessite une intervention humaine est le choix de la taille des fenêtres utilisées. Enfin, la méthode ZERO [15], [14] qui est plus précise, plus rapide et qui ne nécessite aucun paramètre a été élaborée.

L'algorithme ZERO compte le nombre de zéros pour détecter l'origine de la grille dans chaque fenêtre d'une image. En effet, ZERO exploite le fait que la compression JPEG met des valeurs DCT à zéro et identifie la présence d'une grille JPEG lorsqu'un nombre significatif de zéros est observé pour une origine de grille donnée. Chaque pixel peut appartenir à 64 blocs différents qui se chevauchent et vote pour choisir l'origine de la grille du bloc avec le plus de zéros. La méthode est appliquée de trois manières : d'abord globalement pour détecter l'origine de grille globale de l'image. Ensuite, localement pour détecter les grilles étrangères et enfin, à une version alternative de l'image pour détecter les zones où la grille globale est absente.

Les algorithmes développés se distinguent nettement des approches conventionnelles par leur gestion rigoureuse des faux positifs, notamment en ne

détectant aucun événement dans des images n'ayant jamais subi de compression JPEG, là où d'autres méthodes échouent. Cette efficacité repose sur l'adoption de l'approche *a contrario*, un cadre qui, contrairement aux méthodes basées sur le seuillage simple d'un score de confiance, utilise des tests d'hypothèses pour contrôler l'espérance des fausses alarmes en tenant compte du nombre de tests effectués. Cette méthodologie offre une rigueur mathématique permettant des décisions de détection sans interprétation experte et garantit une fiabilité supérieure des résultats.

En collaboration avec deux autres doctorants, nous avons développé une méthodologie et un jeu de données, nommés *Trace* [2], pour évaluer la fiabilité des méthodes de détection de falsification d'images en se focalisant sur leur sensibilité face à des artefacts spécifiques. Notre approche, en plus d'évaluer, révèle les points forts et faibles des méthodes testées. La méthode ZERO se distingue par sa capacité à éviter les faux positifs, affichant d'excellentes performances comparées à 16 autres algorithmes, malgré certaines limitations comme l'omission de petites falsifications ou la délimitation imprécise des zones altérées.

Impact des travaux

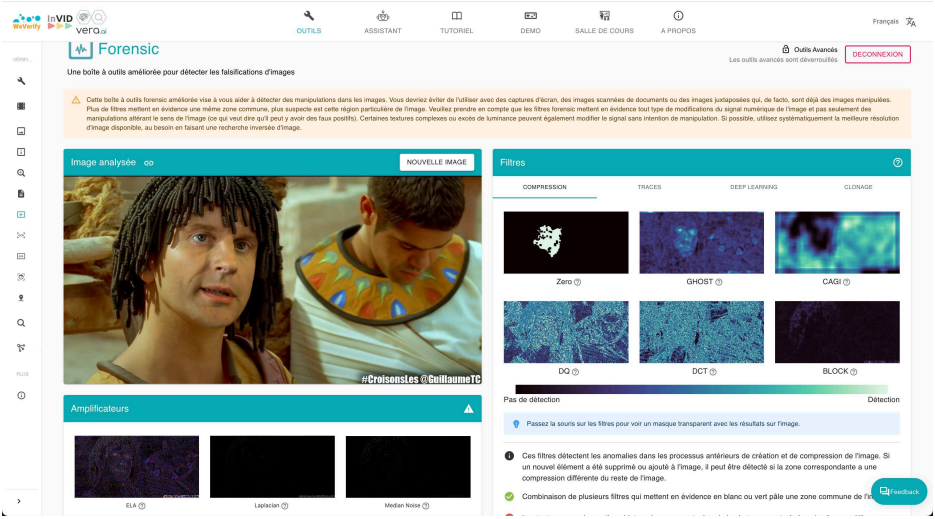


Fig. 3. Capture d'écran de l'outil InVID-WeVerify de l'AFP. La méthode zero [14] se trouve parmi les outils d'analyse d'images (l'image testée provient du compte X de @GuillaumeTC).

Pour assurer la reproductibilité de mes travaux, je publie mes algorithmes sur IPOL (*Image Processing On Line*, ipol.im), une revue de science ouverte. Cette revue accorde une attention particulière aux détails mathématiques et réalise une évaluation stricte par les pairs pour vérifier que le code source correspond aux spécifications de l'article. Chaque article publié dans le journal IPOL détaille la méthode utilisée, présente le pseudo-code, et inclut le code source correspondant. De plus, un démonstrateur en ligne est mis à disposition, offrant la possibilité de tester la méthode avec ses propres images [17], [14], [18].

Cela a permis à ces méthodes d'être largement testées, et en raison de sa rapidité, de sa fiabilité et de sa précision, l'algorithme ZERO a été inclus dans le plugin InVID-WeVerify [9] (comme illustré figure 3). Cet outil, créé par l'Agence France-Presse (AFP), est dédié à la vérification d'informations. Il est accessible à tous et est utilisé principalement par les journalistes et *fact-checkers* dans 224 pays [1]. À travers cet outil, la méthode ZERO a notamment permis de détecter des cas réels de falsification, comme illustré dans cet article de France24 [6].

Par ailleurs, j'ai également adapté les sujets de cette thèse pour différents publics. Ils sont enseignés dans le cadre du cours de traitement des images, faisant partie de la formation en « Éducation aux médias et à l'information » dans le cadre du diplôme universitaire entre l'ENS Paris-Saclay et l'ESJ Lille. Ces idées sont également présentées lors de conférences et diffusées à travers des publications destinées au grand public [12], [13], y compris aux jeunes.

Références

[1] Agence France-Presse. 2023. Le medialab de l'AFP. Retrieved from `https://www.afp.com/fr/lagence/medialab`.

[2] Quentin Bammey, Tina Nikoukhah, Marina Gardella, Rafael Grompone von Gioi, Miguel Colom, and Jean-Michel Morel. 2022. Non-semantic evaluation of image forensics tools: Methodology and database. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 3751–3760.

[3] Davide Cozzolino and Luisa Verdoliva. 2020. Noiseprint: A CNN-based camera model fingerprint. *IEEE Transactions on Information Forensics and Security* 15:144–159. `https://doi.org/10.1109/TIFS.2019.2916364`.

[4] Agnès Desolneux, Lionel Moisan, and Jean-Michel Morel. 2007. *From gestalt theory to image analysis. Interdisciplinary applied mathematics*. Springer, New York. `https://doi.org/10.1023/A:1026593302236`.

[5] Hany Farid. 2009. Exposing digital forgeries from JPEG ghosts. *IEEE transactions on information forensics and security* 4, 1:154–160.

[6] France24. 2023. Une photo de bébé israélien calciné générée par IA : Quand les outils de détection brouillent le débat public. Retrieved from `https://observers.france24.com/fr/moyen-orient/20231017-une-photo-de-b%C3%A9b%C3%A9-isra%C3%A9lien-calcin%C3%A9-g%C3%A9n%C3%A9r%C3%A9e-par-ia-quand-les-outils-de-d%C3%A9tection-brouillent-le-d%C3%A9bat-public`.

[7] Fabrizio Guillaro, Davide Cozzolino, Avneesh Sud, Nicholas Dufour, and Luisa Verdoliva. 2023. TruFor: Leveraging all-round clues for trustworthy image forgery detection and localization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 20606–20615.

[8] Chryssanthi Iakovidou, Markos Zampoglou, Symeon Papadopoulos, and Yiannis Kompatsiaris. 2018. Content-aware detection of JPEG grid inconsistencies for intuitive image forensics. *Journal of Visual Communication and Image Representation* 54:155–170.

[9] InVID-WeVerify. 2023. InVID-WeVerify plugin. Retrieved from `https://weverify.eu/verification-plugin/`.

[10] Neal Krawetz and Hacker Factor Solutions. 2007. A picture's worth. *Hacker Factor Solutions* 6, 2.

[11] Hannes Mareen, Dante Vanden Bussche, Fabrizio Guillaro, Davide Cozzolino, Glenn Van Wallendael, Peter Lambert, and Luisa Verdoliva. 2022. Comprint: Image forgery detection and localization using compression fingerprints. In *International conference on pattern recognition*, 281–299.

[12] Tina Nikoukhah. 2020. Tout ce que les algorithmes de traitement d'images font pour nous. *Interstices*. Retrieved from `https://hal.inria.fr/hal-02929192`.

[13] Tina Nikoukhah. 2021. Les traces de compression pour détecter les photomontages. *Interstices*. Retrieved from `https://hal.inria.fr/hal-03220223`.

[14] Tina Nikoukhah, Jérémy Anger, Miguel Colom, Jean-Michel Morel, and Rafael Grompone von Gioi. 2021. ZERO: a Local JPEG Grid Origin Detector Based on the Number of DCT Zeros and its Applications in Image Forensics. *Image Processing On Line* 11:396–433.

[15] Tina Nikoukhah, Jérémy Anger, Thibaud Ehret, Miguel Colom, Jean-Michel Morel, and Rafael Grompone von Gioi. 2019. JPEG grid detection based on the number of DCT zeros and its application to automatic and localized forgery detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 110–118.

[16] Tina Nikoukhah, Miguel Colom, Jean-Michel Morel, and Rafael Grompone von Gioi. 2019. Détection de grille JPEG par compression simulée. In *XXVIIe colloque sur le traitement du signal et des images*, 641–644.

[17] Tina Nikoukhah, Miguel Colom, Jean-Michel Morel, and Rafael Grompone von Gioi. 2020. Local JPEG Grid Detector via Blocking Artifacts, a Forgery Detection Tool. *Image Processing On Line* 10:24–42.

[18] Tina Nikoukhah, Miguel Colom, Jean-Michel Morel, and Rafael Grompone von Gioi. 2022. A Reliable JPEG Quantization Table Estimator. *Image Processing On Line* 12:173–197.

[19] Tina Nikoukhah, Rafael Grompone von Gioi, Miguel Colom, and Jean-Michel Morel. 2018. Automatic JPEG grid detection with controlled false alarms, and its image forensic applications. In *2018 IEEE conference on multimedia information processing and retrieval (MIPR)*, 378–383.

[20] Yue Wu, Wael AbdAlmageed, and Premkumar Natarajan. 2019. ManTra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features. In *The IEEE conference on computer vision and pattern recognition (CVPR)*.

Cryptanalyse physique à textes inconnus d'algorithmes de chiffrement symétriques

Hélène Le Bouder[1]

La cryptanalyse est une discipline qui tente de briser les solutions cryptographiques. Elle est généralement basée sur une analyse logique et théorique. Dans cet article, ce sont les failles du circuit qui exécute la cryptographie, qui sont utilisées pour attaquer. On parle de cryptanalyse physique. Cet article résume l'approche utilisée dans les publications [11, 17]. Plus précisément, il présente des attaques physiques ciblant des algorithmes de chiffrement symétrique. Ces attaques utilisent la propagation de croyance basée uniquement sur l'observation de mesures, l'attaquant n'a accès ni aux données d'entrée ni à celles de sortie.

Un algorithme cryptographique est conçu pour être mathématiquement robuste. Malheureusement, une fois implémenté dans un circuit, un attaquant peut exploiter les failles de ce dernier. Les attaques physiques sont l'ensemble des techniques exploitant le fonctionnement d'un circuit. Ces attaques nécessitent un accès physique au circuit et sont divisées en deux grandes familles :

– les attaques par observation ;
– les attaques par injection de faute.

1. IMT-Atlantique, OCIF, IRISA, Rennes, France.

DOI:10.48556/SIF.1024.24.159

Cet article traite de la première famille : les attaques par observation. Celles-ci se basent sur l'observation de paramètres physiques pendant l'exécution d'un calcul.

Cet article traite plus spécifiquement d'attaques par observation dites à textes inconnus. C'est-à-dire que l'attaquant n'a accès ni aux données d'entrée ni de sortie. Les cibles choisies sont des algorithmes de chiffrement symétrique.

Dans une première partie, la section « Contexte cryptographique symétrique » aborde le contexte de la cryptographie symétrique. Ensuite, le contexte des attaques par observation est présenté dans la section « Attaques par observation ». Le déroulement des attaques par observation dites à textes inconnus, contre des algorithmes de chiffrement symétrique est décrit dans la section « Attaques sur algorithme de chiffrement symétrique ».

Contexte cryptographique symétrique

Définitions et propriétés

La cryptologie, étymologiquement la science du secret, regroupe la cryptographie et la cryptanalyse.

La cryptographie est la discipline qui assure les propriétés de confidentialité, d'intégrité, d'authentification et de non-répudiation. La cryptanalyse est la discipline qui tente de briser les solutions cryptographiques. La cryptanalyse physique est la discipline qui utilise des attaques physiques pour attaquer des algorithmes cryptographiques.

La cryptographie symétrique est un sous-domaine de la cryptographie. Elle regroupe des algorithmes qui assurent la confidentialité ou l'intégrité, en s'appuyant sur les principes de diffusion et de confusion définis par Shannon [18].

La confusion consiste à rendre la relation entre les données d'entrée et les données de sortie la plus complexe possible. La diffusion repose sur l'idée qu'un biais dans les données d'entrée ne doit pas se retrouver dans les données de sortie. Autrement dit, la connaissance de la sortie doit fournir le moins d'information possible sur les données d'entrée.

Chiffrement symétrique

Le chiffrement est l'un des sujets historiques de la cryptographie. Il s'agit de l'ensemble des méthodes permettant de rendre un message, appelé texte clair noté M, illisible à toute personne autre que le destinataire. Il ne s'agit pas de dissimuler le message, ce qui relève de la stéganographie, mais de le chiffrer à l'aide d'une clé notée K, afin de le transformer en quelque chose d'incompréhensible pour quiconque ne possède pas la clé, appelé texte chiffré, noté C. La méthode utilisée est appelée algorithme de chiffrement.

$$C = En(M, K) \tag{1}$$

L'opération inverse est appelée algorithme de déchiffrement.

$$M = Dn(C, K). \tag{2}$$

En cryptographie symétrique, la même clé est utilisée pour chiffrer et déchiffrer.

Une méthode classique pour réaliser un chiffrement symétrique est le chiffrement par blocs. Il utilise une primitive bloc prenant en entrée un bloc de taille fixe et une clé. Cette primitive retourne une sortie de même taille que le bloc d'entrée. Ensuite, une autre fonction, appelée mode de chiffrement, combine les différents blocs. La plupart des modes nécessitent de plus un vecteur d'initialisation, pour chaque opération de chiffrement.

Code d'authentification de message

Un MAC[2] est un outil de cryptographie symétrique qui assure l'intégrité. Une fonction MAC prend en entrée une clé secrète K et un message de longueur quelconque M. Elle retourne un tag de longueur fixe T.
Pour envoyer un message M protégé en intégrité, on calcule le tag :

$$T = \mathrm{MAC}(M, K) \tag{3}$$

Pour vérifier l'intégrité, on calcule $\mathrm{MAC}(M, K)$ et on teste si cela est égal à T :

– si oui, on considère que l'intégrité est respectée ;
– sinon, on considère que l'intégrité n'est pas respectée.

Un MAC peut se calculer à l'aide d'une primitive bloc et d'un mode MAC.

Chiffrement intègre

Bien souvent, lors d'un échange, la confidentialité et l'intégrité sont nécessaires. Pour garantir ces deux propriétés, plusieurs solutions symétriques sont proposées dans la littérature :

– une première solution est l'utilisation de deux outils. On combine un chiffrement et un MAC ;
– une autre solution est d'utiliser un chiffrement par blocs avec un mode garantissant les deux propriétés, tel que GCM[3] ;
– enfin, une dernière solution, proposée récemment, est d'adopter un nouvel outil : un chiffrement intègre dit tout en un.

2. *Message authentification code*.
3. *Galois/Counter mode*.

Pour développer cette solution dans un contexte de cryptographie à bas coût, le NIST[4] a lancé une compétition [15] pour établir des standards de chiffrement symétrique qui assurent également l'intégrité des données. En février 2023, il a élu Ascon [6] comme nouveau standard. Les candidats de la compétition NIST appelés en anglais AEAD[5] (attention, le mot « authenticated », qui est parfois traduit par authentifié, signifie ici intègre.), suivent le schéma illustré en Figure 1.

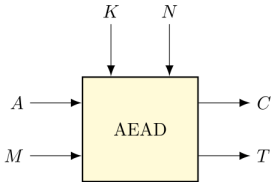


Fig. 1. Schéma générique d'un chiffrement symétrique intégré AEAD.

Ils prennent en entrée un message clair M, des données dites associées A, une clé secrète K et un vecteur d'initialisation N appelé « nonce ». Attention, N doit changer pour chaque message. Ils retournent un tag T, qui permet de vérifier l'intégrité des données, ainsi qu'un texte chiffré C.

Algorithmes cibles

Dans cet article nous présentons deux algorithmes.

L'ALGORITHME AES

L'algorithme AES[6] [14] est une primitive bloc, approuvée par le NIST en 2001, illustrée en figure 2. Il existe sous trois formes : 10, 12 ou 14 tours, avec des clés respectives de 128, 192 et 256 bits ; dans cet article, nous utilisons la version 128 bits. Et l'AES traite des blocs de 16 octets vus comme une matrice 4×4 appelée « state ». La fonction de tour de l'AES se décompose en les quatre sous-fonctions ci-dessous.

- **SubBytes** est composé de 16 s-boxes identiques. Cette fonction s-box, notée SB, est une fonction booléenne de 8 bits vers 8 qui s'applique à tous les octets du state.
- **ShiftRows**, noté SR, déplace cycliquement vers la gauche les trois dernières lignes du state de respectivement un, deux et trois éléments.
- **MixColumns**, noté MC, applique une transformation linéaire aux colonnes du state.

4. *National Institute of Standards and Technology*.
5. *Authenticated encryption with associated data*.
6. *Advanced encryption standard*.

– **AddRoundKey** effectue une opération « xor » avec la clé de tour.

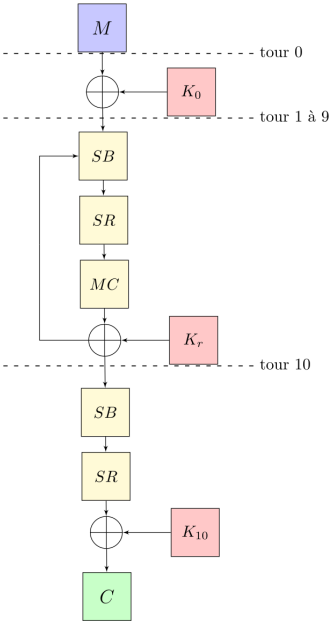


Fig. 2. Schéma de l'AES de 10 tours.

Les fonctions *SR* et *MC* servent à la diffusion ; un bit changé sur un octet affecte l'ensemble du state après deux tours. Il est important de préciser que le tour 0 de l'AES effectue uniquement un xor avec la première clé, et que le dernier tour ne contient pas de *MC*.

Une clé de tour de l'AES est représentée comme le state sous la forme d'une matrice 4×4. Ainsi, l'emplacement de chacun des octets est noté par un couple (l, c), représentant respectivement la ligne l et la colonne c dans la matrice.

Pour un AES_{128}, la clé maître K est la clé du tour 0, K_0. La clé de tour K_{r+1} est obtenue à partir de la clé du tour précédent K_r, en utilisant les équations (4) suivantes :

$$\begin{cases} K_{r+1}^{l,0} &= SB\left(K_r^{(l+1) \bmod 4,3}\right) \oplus rcon(l,r) \oplus K_r^{l,0} & \forall l \in [\![0,3]\!] \\ K_{r+1}^{l,c} &= K_r^{l,c} \oplus K_{r+1}^{l,c-1} & \forall l \in [\![0,3]\!] \text{ et } c \in [\![1,3]\!] \end{cases} \tag{4}$$

avec SB la fonction s-box, et $rcon$ désigne une matrice constante de taille 4×10.

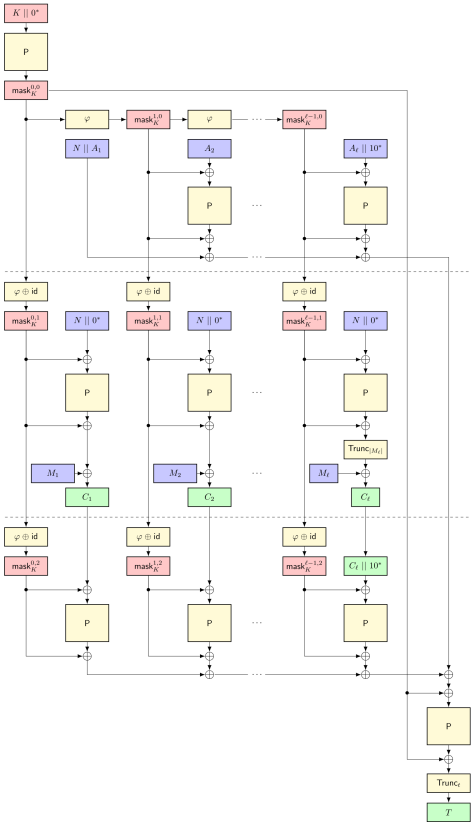


Fig. 3. L'initialisation est tout à gauche, suivie du chiffrement. La génération du tag est tout à droite.

Elephant [2] est un finaliste AEAD de la compétition de cryptographie légère du NIST. Soit P une permutation cryptographique de n bits, et φ un LFSR[7] de n bits. Soit la fonction $\text{mask} : [\![0,1]\!]^{128} \times \mathbb{N} \times \{0,1,2\} \to [\![0,1]\!]^n$ telle que :

$$\text{mask}_K^{i,j} = \text{mask}(K,i,j) = (\varphi \oplus \text{id})^j \circ \varphi^i \circ \text{P}\left(K||0^{n-128}\right), \tag{5}$$

où id est la fonction identité.

Le chiffrement En prend en entrée une clé de 128 bits K, un nonce de 96 bits N, des données associées A et un texte en clair M. Il produit un texte chiffré C

7. *Linear-feedback shift register*, registre à décalage à rétroaction linéaire.

de la même taille que M, et un tag T de ν bits. Le schéma est illustré sur la figure 3.

Le déchiffrement prend en entrée la clé de 128 bits K, le nonce de 96 bits N, les données associées A, le texte chiffré C, et le tag T de ν bits. Il retourne un texte en clair M de la même taille que C si et seulement si le tag T est correct ; sinon, il retourne un message d'erreur.

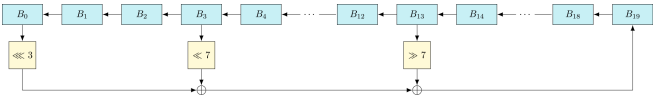


Fig. 4. 160-bit LFSR φ_{Dumbo}

Elephant se décline en trois versions qui diffèrent par la permutation cryptographique de n bits P et le LFSR φ utilisés, ainsi que par la taille du tag ν. Dans cet article, nous n'utiliserons que la version Dumbo. Dumbo utilise une permutation de 160 bits [3], le LFSR φ_{Dumbo} donné par l'équation (6), illustré sur la Figure 4, et a une taille de tag $\nu = 64$ bits.

$$\varphi_{\text{Dumbo}} : (B_0, \cdots, B_{19}) \mapsto (B_1, \cdots, B_{19}, (B_0 \lll 3) \oplus (B_3 \ll 7) \oplus (B_{13} \gg 7)) \tag{6}$$

Attaques par observation

Entre l'algorithme conçu de manière théorique et sa mise en pratique dans un circuit (appelée aussi implémentation logicielle ou matérielle), il y a différents niveaux d'abstraction. À chaque niveau d'abstraction, de nouvelles failles sont possibles, c'est pourquoi les algorithmes théoriques ne peuvent pas anticiper toutes les attaques (présentes et futures) à la fois. Voici une liste des différents niveaux.

1. **Algorithme :** la représentation abstraite.
2. **Logiciel :** le langage de programmation. Une fois conçu, l'algorithme est décrit dans un langage de programmation. Le programme est alors compilé afin d'être traduit en langage machine, sous forme d'un ensemble d'instructions.
3. **Architecture de la machine et micro-architecture :** le processeur ou CPU est le composant de l'ordinateur qui exécute les instructions machines. Un processeur construit en un seul circuit intégré est appelé microprocesseur.
4. **Circuits logiques :** une instruction de type ALU se décompose en fonctions booléennes. Elles sont mises en œuvre en électronique sous forme de portes logiques. Une porte logique réalise des opérations booléennes sur une séquence de bits.

5. **Transistors :** dispositifs semi-conducteurs qui permettent de contrôler un courant (ou une tension). Les portes logiques sont construites à partir de plusieurs transistors connectés de manière adéquate. Dans un transistor, un 0 logique correspond à une tension en dessous d'une certaine valeur, sinon cela correspond à un 1 logique.

Si un algorithme est théoriquement *sûr*, le circuit dans lequel il est implémenté ne l'est pas forcément. C'est la frontière entre la théorie et la pratique. L'algorithme est théorique, le circuit, lui, est bien concret. Ainsi, ce dernier a une consommation de courant, un temps de calcul, etc. Ce sont ces paramètres physiques qui peuvent nuire à la sécurité de notre algorithme.

Les attaques par observation, appelées aussi parfois attaques par canaux auxiliaires, notées SCA[8], s'effectuent aux niveaux circuit logique et transistors et ciblent généralement les niveaux logiciel et algorithme.

On observe depuis quelques années une augmentation de l'utilisation de petits circuits, connectés ou non, dans la vie quotidienne. Or, les systèmes embarqués et autres petits circuits sont des cibles privilégiées des attaques par observation. Le but de ces attaques est généralement :
– la recherche d'un secret, comme une clé cryptographique, ce qui est le cas dans la suite de cet article ;
– ou la rétroconception d'algorithmes.

Fuites physiques

Une attaque par observation analyse une fuite physique d'un circuit. Il s'agit de mesures physiques comme le temps, la température, la consommation de courant ou le rayonnement électromagnétique. Dans cet article, nous nous focalisons sur les deux derniers.

CONSOMMATION DE COURANT

L'idée est de se brancher directement sur le circuit pour mesurer sa consommation de courant. Un courant électrique est un déplacement de charges électriques. Chaque instruction réalisée par un circuit met en œuvre un certain nombre de transistors. À chaque instant, la mesure de l'intensité du courant électrique, appelée consommation de courant, reflète l'activité du circuit.

RAYONNEMENT ÉLECTROMAGNÉTIQUE

Les émissions d'ondes électromagnétiques d'un circuit sont causées par la circulation de courant en son sein. Cette circulation de charges dans le conducteur produit un champ électromagnétique composé d'un champ magnétique et d'un champ électrique, mathématiquement reliés par les équations de Maxwell. Ces ondes électromagnétiques se propagent dans l'espace et peuvent ainsi être

8. *Side channel analysis*.

captées par une sonde. La sonde doit généralement être positionnée aussi près que possible de la surface active du circuit ; on parle de mesures en champ proche. La figure 5 représente une plateforme permettant d'acquérir le rayonnement électromagnétique d'un circuit. Cette plateforme est la propriété de l'INRIA.

Fig. 5. EMA, plateforme de mesure de rayonnement électromagnétique du Laboratoire haute sécurité (LHS) d'Inria Rennes.

L'avantage majeur d'une analyse électromagnétique (EM) par rapport à une analyse par consommation de courant est que les mesures sont prises localement. Un autre avantage est que ces attaques sont non-invasives. Le circuit ne peut pas facilement détecter qu'il est observé, car il n'y a ni branchement ni même un contact physique. Cependant, les mesures EM sont souvent plus sensibles au bruit et dépendent fortement du positionnement de la sonde. La consommation de courant et l'EM sont étroitement liées, elles sont donc souvent modélisées et utilisées de manière très similaire.

Modèles mathématiques

Un circuit intégré est constitué d'un ensemble de portes logiques. La consommation de courant du circuit est la somme des consommations de courant de chacune de ses portes logiques. Une porte logique n'a pas une consommation de

courant constante dans le temps. Il existe deux modèles de consommation de courant pour les portes logiques :
– un modèle dynamique où une porte logique consomme de l'énergie à chaque changement d'état ;
– un modèle statique où une porte logique consomme de l'énergie même au repos.

Ces deux modèles sont couramment utilisés. On peut donc distinguer une transition de 0 vers 1 d'une transition de 1 vers 0. C'est pourquoi la consommation de courant d'une valeur présente dans le circuit, portée par un ensemble de portes logiques, est considérée comme proportionnelle :
– à son HW[9], c'est-à-dire le nombre de bits à 1 ;
– ou à sa HD[10] avec la dernière valeur calculée, c'est-à-dire le nombre de variations de bits entre deux valeurs.

Un attaquant qui récupère le HW d'un octet obtient de l'information sur celui-ci. En effet, un octet B peut prendre 256 valeurs dans $[\![0, 255]\!]$. Connaissant le HW de B, l'attaquant réduit la liste des valeurs possibles, comme le montre le tableau 1.

$\mathrm{HW}(B)$	0	1	2	3	4	5	6	7	8
$\#B$	1	8	28	56	70	56	28	8	1

Table 1. Nombre de valeurs possibles pour un octet B selon son poids de Hamming HW.

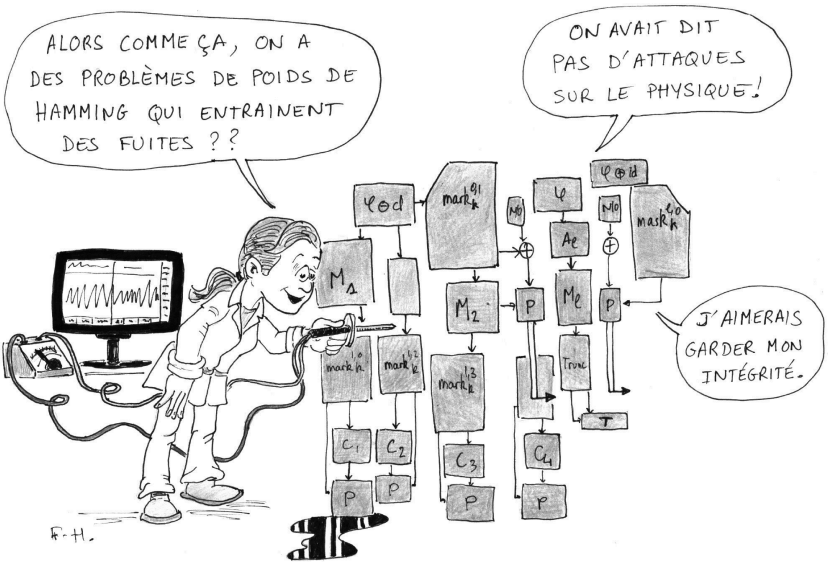
Si le poids de Hamming est souvent utilisé comme modèle représentant la consommation de courant ou le rayonnement électromagnétique, un modèle plus précis est un poids de Hamming bruité $\widetilde{\mathrm{HW}}$. C'est le modèle utilisé dans cet article. Soit B un octet, le poids de Hamming bruité $\widetilde{\mathrm{HW}}$ se calcule comme suit :

$$\widetilde{\mathrm{HW}}(B) = \mathrm{HW}(B) + \sigma_{B,t} \tag{7}$$

où $\sigma_{B,t}$ un événement de la variable aléatoire gaussienne $\mathcal{N}\left(0, \sigma^2\right)$ à un instant t . La fonction de densité de probabilité $\mathbf{F}$ associée à $\mathcal{N}\left(0, \sigma^2\right)$ est donnée par :

$$\mathbf{F}_{\sigma}(B) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\frac{B}{\sigma}\right)^2\right) \tag{8}$$

9. Hamming weight.
10. Hamming distance.



Attaques sur algorithme de chiffrement symétrique

Modèle de l'attaquant

Dans cet article, nous présentons des attaques à textes inconnus. Cela signifie que l'attaquant ne possède ni les données d'entrée ni de sortie de l'algorithme ciblé. Il a connaissance uniquement des traces mesurées, modélisées par des poids de Hamming bruités sur les octets. Ces attaques sont notées BSCA[11].

Les deux premières publications [11, 12] de ce type sont apparues à la même période, dans le but d'attaquer l'AES. Ces travaux ont été améliorés par Christophe Clavier *et al.* dans [4]. C'est dans cette contribution qu'apparaît pour la première fois le terme BSCA. Aujourd'hui, c'est une nouvelle famille d'attaques par observation [5].

Les attaques BSCA n'ont pas encore prouvé leur faisabilité en pratique. Elles travaillent avec des simulations de consommation de courant, généralement un poids de Hamming bruité. Néanmoins, cette nouvelle famille d'attaques est un très bon outil pour tester des primitives cryptographiques symétriques.

Algorithme de propagation de croyance

Dans la suite de cet article, nous supposons que l'attaquant a récupéré des poids de Hamming bruités. Il va alors définir un chemin d'attaque, c'est-à-dire relier

11. *Blind side channel analysis*.

ses mesures au secret attaqué (une clé) via les calculs de l'algorithme. Puis, il regroupe toutes ces informations pour obtenir le secret.

À cette fin, une technique connue sous le nom d'algorithme de propagation de croyance, BP[12] [10], est utilisée. BP est une approche d'inférence bayésienne [1]. Il a été utilisé pour la première fois par Gallager [7, 8] pour le décodage des codes correcteurs. Il a ensuite été redécouvert par Tanner [19] et formalisé par Pearl [16]. La première fois que BP a été utilisé en SCA, c'était dans l'attaque [20], puis il a été étudié dans [9]. Enfin, cet outil a été utilisé pour des attaques BSCA [11, 13].

GRAPHE DE TANNER

Un algorithme BP repose sur un graphe de Tanner. Les nœuds d'un graphe factoriel sont de deux types :

– des nœuds variables B (notés B car nous travaillons sur des octets) ;
– des nœuds d'équations $\mathcal{E}$.

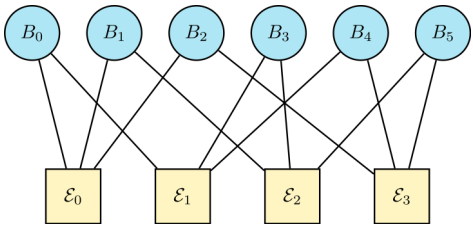


Fig. 6. Exemple d'une partie d'un graphe de Tanner. Les cercles sont des nœuds variables et les carrés sont des nœuds d'équations.

Une arête relie un nœud de variable B_i à un nœud d'équation $\mathcal{E}_j$, lorsque l'équation $\mathcal{E}_j$ représentée par le nœud d'équation implique la variable B_i. Un extrait d'un graphe de Tanner est illustré dans la Figure 6.

N désigne l'ensemble des voisins d'un nœud. Ainsi, l'ensemble des $\mathrm{N}(\mathcal{E})$ est composé de nœuds variables B, et l'ensemble des $\mathrm{N}(x)$ est composé de nœuds d'équations $\mathcal{E}$.

PROPAGATION DE LA CROYANCE

Les entrées d'un algorithme de propagation des croyances (BP) sont :

– un graphe de Tanner ;
– des probabilités a priori $\mathbb{P}_A(B_i = b)$ sur les différents nœuds de variables B.

12. *Belief propagation*.

L'algorithme BP retourne des probabilités, à partir des valeurs initiales $\mathbb{P}_A(B_i = b)$. Ces probabilités sont dites a posteriori $\mathbb{P}_P(B_i = b)$ et portent sur les différents nœuds de variables.

L'idée est que l'information se diffuse à travers le graphe de Tanner. Les nœuds voisins s'échangent de l'information et la partagent avec les autres voisins, et ainsi de suite.

En pratique, les nœuds dans le graphe échangent des messages avec leurs voisins. Plus précisément, puisque le graphe est biparti, deux types de messages sont échangés :

– des messages d'un nœud d'équation E vers un nœud de variable V, notés $\mu_{E \to V}$;
– des messages d'un nœud de variable V vers un nœud d'équation E, notés $\mu_{V \to E}$.

Les messages de variable à équation $\mu_{V \to E}(v)$ sont initialisés avec les probabilités a priori $\mathbb{P}_A(B_i = b)$.

Puis, BP fonctionne en appliquant alternativement les messages de chaque type pour mettre à jour les probabilités. À la fin de l'exécution, la valeur retournée est appelée probabilité a posteriori. Le nombre d'itérations dépend de la rapidité de convergence, qui elle-même dépend du graphe de Tanner.

Application sur différents algorithmes

ATTAQUE SUR AES

Dans cette partie, l'attaque BSCA sur AES [11] est résumée.

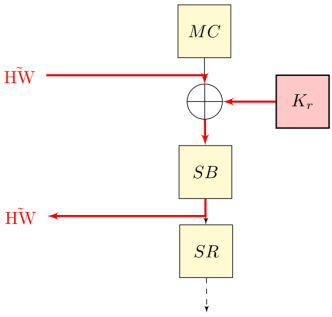


Fig. 7. Chemin d'attaque sur l'AES

Chemin d'attaque

L'acquisition de traces se déroule à chaque tour de l'AES, exceptés 0 et 10. L'objectif est de récupérer le poids de Hamming bruité des différents octets du state

avant le xor de clé et après le **SubBytes**. Il n'y a pas de **MixColumns** avant le « xor » avec K_0, et il n'y a pas de **SubBytes** après K_{10}.

Ensuite, une première étape consiste, à partir des mesures de ces poids de Hamming bruités, à attribuer une probabilité à chacun des octets des différentes clés de tours.

Graphe de Tanner

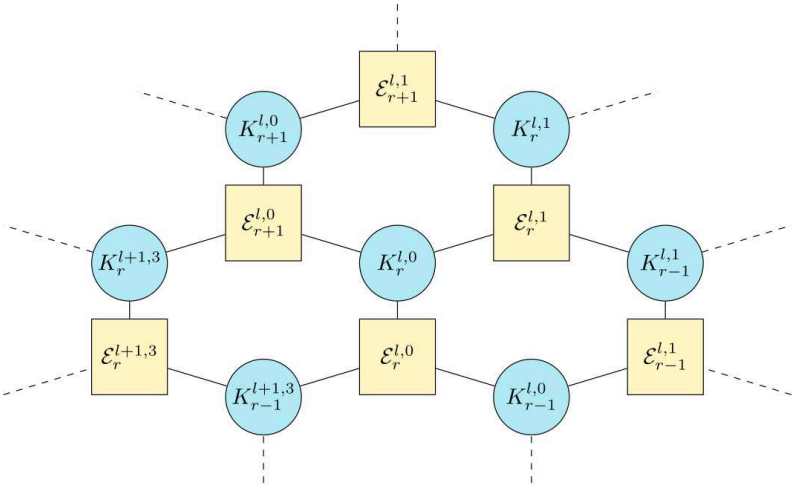


Fig. 8. Extrait du graphe de Tanner basé sur les équations liant les clés de tours de l'AES.

Dans l'AES, les différentes clés de tour sont dérivées de la clé secrète utilisée au premier tour. De plus, l'attaquant dispose de mesures permettant d'attribuer une probabilité à chacun des octets des différentes clés de tours. Ainsi, le graphe de Tanner est construit sur l'algorithme de dérivation des clés de tours. Il est illustré en Figure 8. Les équations $\mathcal{E}$ définies en #eq:KeyExpansion relient les différents octets des clés de tours.

Résultats

Les résultats illustrés dans le tableau 2 montrent que l'attaque est efficace. L'utilisation de l'algorithme BP est un excellent moyen de récupérer la clé. Même en présence d'un bruit élevé, l'attaque peut réussir. L'approche montre qu'augmenter le nombre de tours dans une primitive bloc, afin de le rendre résistant à la cryptanalyse classique, peut l'affaiblir dans notre cas.

$n \backslash \sigma$	0.1	0.2	0.3	0.5	1.0	1.5	2.0	3.0
100	✓	✓	✓	✓	×	×	×	×
1000	✓	✓	✓	✓	✓	×	×	×
10000	✓	✓	✓	✓	✓	✓	✓	×
100000	✓	✓	✓	✓	✓	✓	✓	✓

Table 2. Succès de l'attaque en fonction de l'écart-type du bruit σ et du nombre de traces n, pour 100 attaques simulées. ✓ indique que l'attaque a réussi et × indique que l'attaque a échoué.

ATTAQUE SUR ELEPHANT

Dans ce paragraphe, un autre exemple d'application de ce type d'attaque est présenté.

Chemin d'attaque

Les LFSR sont souvent utilisés en cryptographie légère, et leurs états initiaux dépendent souvent à la fois de la clé K et du nonce N. Étant donné que le nonce doit être modifié pour chaque chiffrement, les attaques sur de tels schémas sont limitées à l'algorithme de déchiffrement. Dans le cas d'Elephant, le LFSR dépend uniquement de la clé secrète. Par conséquent, notre attaque peut être appliquée dans un scénario de chiffrement. L'objectif de l'attaque est alors de récupérer l'état initial secret du LFSR d'Elephant. Récupérer l'état initial du LFSR équivaut à récupérer la clé secrète. En effet, l'état initial n'est que le résultat de la permutation connue P appliquée à la clé. Le LFSR générant un nouvel octet à chaque itération, le contenu du φ_{Dumbo} LFSR est noté comme suit :

$$(B_j, \cdots, B_{j+19}) = \mathrm{mask}_K^{j,0} \tag{9}$$

et soit B_{j+20} l'octet généré à l'itération j. De même, soit $(B_j^1, \cdots, B_{j+19}^1) = \mathrm{mask}_K^{j,1}$ et $(B_j^2, \cdots, B_{j+19}^2) = \mathrm{mask}_K^{j,2}$. Par définition du $mask$, les relations suivantes sont valables pour tout $j \geq 0$:

$$B_j^1 = B_j \oplus B_{j+1} \tag{10}$$

$$B_j^2 = B_j \oplus B_{j+2} \tag{11}$$

L'attaquant peut exploiter deux vecteurs d'attaque : d'une part, les équations (6) provenant de l'itération du LFSR, et d'autre part, les équations (10) et (11) provenant des différents masques utilisés pour la séparation des domaines. Dans un premier temps, l'attaquant obtient un HW bruité sur les différents octets des φ_{Dumbo} LFSR.

Graphe de Tanner

Dans le cas d'une attaque sur Elephant, le graphe de Tanner est construit comme suit. Les nœuds de variables sont les octets B_i, B_j^1 et B_k^2 des LFSR. Les nœuds de facteurs représentent :
– les équations de type (10) ;
– les équations de type (11) ;
– les équations de rétroaction du LFSR (6).

Le graphe de Tanner d'Elephant est illustré dans la Figure 9.

Résultats

Pour 1000 clés générées aléatoirement et 11 valeurs de bruit différentes dans $0.1 \leq \sigma \leq 1$, la moyenne, l'écart type, la médiane et les quartiles des rangs des 20 octets de clé corrects ont été calculés. Les résultats sont donnés dans 3.

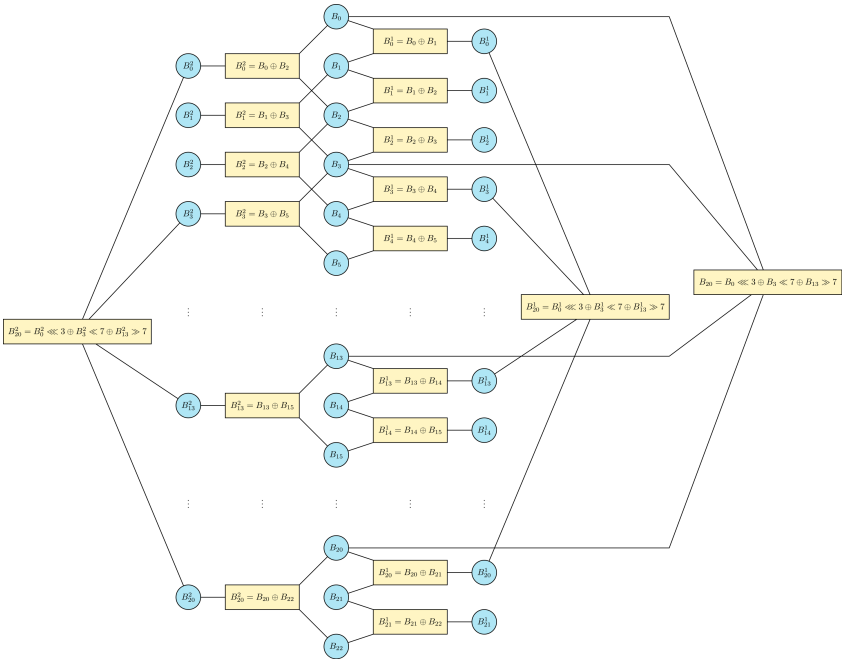


Fig. 9. Extrait du graphe de Tanner du LFSR φ_{Dumbo} d'Elephant.

σ	Moyenne	Écart type	Quartile Q1	Médiane	Quartile Q3	Max
0.1	3.54	5.97	0	3	3	27
0.2	3.67	6.11	0	3	3	31
0.3	6.00	10.76	0	3	3	97
0.4	9.59	15.72	0	3	8	97
0.5	15.97	23.03	3	8	31	153
0.6	23.65	31.41	3	8	31	157
0.7	33.73	40.11	3	27	36	213
1	70.68	61.42	31	36	92	246

Table 3. Rang des 20 octets de la clé pour 1000 clés générées aléatoirement et selon différents bruits σ.

Les résultats montrent que lorsque le niveau de bruit est faible, BP remonte les bonnes hypothèses d'octets de clé dans les premiers rangs. Un attaquant peut alors exploiter les probabilités a posteriori renvoyées par BP et utiliser des techniques d'énumération de clés pour retrouver la clé entière. Notons que, puisque le calcul du LFSR ne dépend que de la clé, il est possible de réduire le niveau de bruit en moyennant plusieurs traces.

La conception d'Elephant, où il existe des relations entre les différents masques du LFSR, constitue une vraie vulnérabilité face à cette attaque.

Conclusion

Dans cet article, le principe de la cryptanalyse physique à textes inconnus est présenté et illustré sur deux algorithmes de chiffrement symétrique. Ces travaux montrent que la cryptanalyse physique, qui utilise l'environnement physique d'un circuit pour en extraire des secrets, peut être employée avec des modèles d'attaquant ayant accès à peu d'informations. Il est ici possible de retrouver le secret uniquement à partir de la trace d'exécution du chiffrement, sans connaître ni les entrées ni les sorties. Cette approche consiste à dérouler un algorithme de propagation de croyance sur l'algorithme attaqué. Bien qu'elle en soit encore au stade théorique, elle permet d'anticiper les attaques pratiques.

Références

[1] David Barber. 2011. *Bayesian Reasoning and Machine Learning*. Cambridge University Press.

[2] Tim Beyne, Yu Long Chen, Christoph Dobraunig, and Bart Mennink. 2021. Elephant v2. *NIST lightweight competition*.

[3] Andrey Bogdanov, Miroslav Knezevic, Gregor Leander, Deniz Toz, Kerem Varici, and Ingrid Verbauwhede. 2011. Spongent: a Lightweight Hash Function. In *CHES*.

[4] Christophe Clavier and Léo Reynaud. 2017. Improved blind side-channel analysis by exploitation of joint distributions of leakages. In *CHES*.

[5] Christophe Clavier, Léo Reynaud, and Antoine Wurcker. 2018. Quadrivariate improved blind side-channel analysis on boolean masked AES. In *COSADE*.

[6] Christoph Dobraunig, Maria Eichlseder, Florian Mendel, and Martin Schläffer. 2014. Ascon. *Submission to the CAESAR competition*.

[7] Robert G. Gallager. 1962. Low-density parity-check codes. *IRE Trans. on Information Theory*.

[8] Robert G. Gallager. 1963. Low Density Parity check codes. MIT, Cambridge, MA.

[9] Vincent Grosso and François-Xavier Standaert. ASCA, SASCA and DPA with Enumeration: Which One Beats the Other and When? In *ASIACRYPT 2015*. Springer.

[10] Frank R. Kschischang, Brendan J. Frey, and Hans-Andrea Loeliger. 2001. Factor graphs and the sum-product algorithm. *IEEE Trans. on Information Theory*.

[11] Hélène Le Bouder, Ronan Lashermes, Yanis Linge, Gaël Thomas, and Jean-Yves Zie. 2016. A Multi-round Side Channel Attack on AES Using Belief Propagation. In *FPS*.

[12] Yanis Linge, Cécile Dumas, and Sophie Lambert-Lacroix. 2014. Using the joint distributions of a cryptographic function in side channel analysis. In *COSADE*.

[13] Julien Maillard, Awaleh Houssein Meraneh, Modou Sarry, Christophe Clavier, Hélène Le Bouder, and Gaël Thomas. 2023. Blind side channel analysis on the Elephant LFSR Extended version. *SECRYPT BOOK*.

[14] NIST. 2001. Specification for the Advanced Encryption Standard. *FIPS PUB 197*.

[15] NIST. 2018. *Lightweight cryptography, standardization process*.

[16] Judea Pearl. 1982. Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach. In *AAAI*.

[17] Modou Sarry, Hélène Le Bouder, Eid Maaloouf, and Gaël Thomas. 2023. Blind side channel analysis against AEAD with a belief propagation approach. In *CARDIS*.

[18] Claude E. Shannon. 1949. Communication Theory of Secrecy Systems. *The Bell System Technical Journal* 28, 4: 656–715.

[19] Robert M. Tanner. 1981. A recursive approach to low complexity codes. *IEEE Trans. on Information Theory*.

[20] Nicolas Veyrat-Charvillon, Benoît Gérard, and François-Xavier Standaert. Soft Analytical Side-Channel Attacks. In *ASIACRYPT 2014*.

KGPRUNE : une application Web pour extraire des sous-graphes d'intérêt de Wikidata par élagage analogique

Pierre Monnin[1], Cherif-Hassan Nousradine[2], Lucas Jarnac[2,3], Laurel Zuckerman[4], Miguel Couceiro[2,5]

Les graphes de connaissance (GC, en anglais *knowledge graphs* ou KG) sont des représentations structurées qui modélisent les connaissances d'un ou plusieurs domaines. Leurs unités atomiques sont des triplets (h, p, o) qui représentent l'existence d'une relation p entre deux entités h et o.

De par leur flexibilité et les connaissances qu'ils fournissent, les GC sont devenus des atouts majeurs et alimentent diverses méthodes d'intelligence artificielle (par exemple, *retrieval augmented generation* pour les grands modèles de langages [19], machine learning en général [4]) avec des applications dans un large éventail de domaines (recherche, e-commerce, réseaux sociaux, sciences de la vie, etc. [3, 9, 18]).

Suivant un des principes fondateurs du Web sémantique [1], les GC existants sont souvent réutilisés pour créer de nouveaux GC ou pour soutenir de nouvelles tâches ou applications [5, 7, 15, 22]. Cette réutilisation est rendue possible par

1. Université Côte d'Azur, Inria, CNRS, I3S, Sophia-Antipolis, France, pierre.monnin@inria.fr.
2. Université de Lorraine, CNRS, LORIA, Nancy, France.
3. Orange, France.
4. Chercheur indépendant.
5. Universidade de Lisboa, INESC-ID, Instituto Superior Técnico, Lisboa, Portugal.

DOI:10.48556/SIF.1024.24.177

le nombre et la taille croissants des GC publiquement accessibles[6], et en particulier des grands GC couvrant plusieurs domaines tels que Wikidata. Ce dernier est un GC générique de plus de 100 millions de nœuds[7] qui soutient Wikipédia [24]. Wikidata est considéré comme une source de connaissances de premier ordre, mais plusieurs problèmes entravent sa réutilisation [11, 21]. Premièrement, sa grande taille entraîne des problèmes de scalabilité lors de la manipulation du graphe (stockage, performances des requêtes, etc.). Deuxièmement, toutes les connaissances représentées ne sont pas pertinentes pour chaque tâche ou application considérée. Par exemple, l'une des entités voisines de `Microsoft SharePoint` est `Dating App` qui peut ne pas être intéressante lors de la création d'un GC d'entreprise modélisant le domaine de l'informatique d'entreprise.

Pour résoudre ces problèmes, plusieurs auteurs ont proposé d'extraire des sous-graphes, soit manuellement avec les premiers exemples remontant à 1996 [22] soit automatiquement [10, 11, 21]. En particulier, nous avons récemment proposé une approche qui parcourt le voisinage d'entités de départ fournies par les utilisateurs, en conservant les voisins pertinents tout en élaguant ceux qui ne le sont pas [10]. Cette approche s'appuie sur l'inférence analogique et présente des performances élevées, y compris dans des situations de transfert, avec un nombre de paramètres particulièrement faible.

En nous appuyant sur ce travail précédent, nous proposons KGPrune,[8] une application Web permettant d'extraire des sous-graphes de Wikidata à partir d'entités de départ et de propriétés intéressantes pour l'utilisateur. KGPrune peut être utilisée à la fois depuis un navigateur et via une API, permettant aux utilisateurs avec diverses expertises techniques d'interagir avec notre approche d'élagage. Dans ce qui suit, après avoir décrit les fonctionnalités et l'architecture technique de KGPrune, nous illustrons son intérêt avec deux applications concrètes : l'amorçage d'un graphe de connaissances d'entreprise et l'extraction de connaissances liées aux œuvres d'art pillées. Une vidéo de démonstration est disponible sur YouTube.[9]

KGPrune : extraire des sous-graphes d'intérêt

Les principaux écrans de l'application Web KGPrune sont présentés en figure 1. Nous décrivons ci-dessous les principales caractéristiques et étapes d'interaction avec l'application.

6. `https://lod-cloud.net/`.
7. `https://www.wikidata.org/wiki/Wikidata:Statistics`.
8. `https://kgprune.loria.fr`.
9. `https://youtu.be/mt5gF4ZmhGY`.

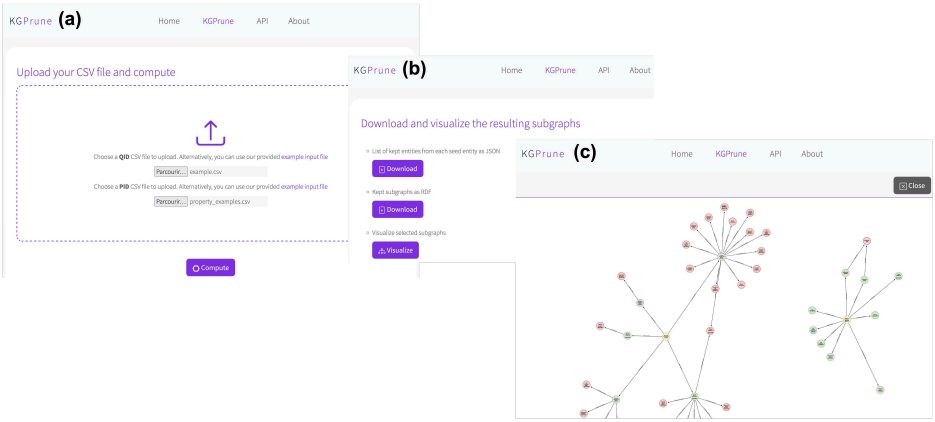


Fig. 1. Écrans principaux de KGPrune. (a) Formulaire de soumission des deux fichiers CSV demandés à l'utilisateur : l'un indiquant les QIDs des entités de départ et l'autre indiquant les PIDs des propriétés à parcourir. (b) Page d'affichage des résultats où l'utilisateur peut choisir de visualiser les sous-graphes extraits ou de les télécharger en JSON ou RDF. (c) Visualisation des sous-graphes extraits où les entités de départ sont en jaune, les voisins conservés en vert et les voisins élagués en rouge.

GC supporté.

Nous avons choisi de construire KGPrune sur Wikidata étant donné sa taille et sa généricité. Il peut donc servir de source de connaissances de premier ordre pour plusieurs domaines. Il convient toutefois de noter que notre approche pourrait être appliquée à n'importe quel GC.

Fichiers en entrée.

KGPrune ne nécessite que deux fichiers CSV comme entrées de l'utilisateur, comme illustré dans la table 1. Le fichier `qid_example.csv` contient des QIDs identifiant les entités de départ d'intérêt dont le voisinage sera récupéré. Ici, à titre d'exemple, nous considérons `Microsoft SharePoint` (Q18833) et le langage de programmation `Java` (Q251). Le fichier `pid_example.csv` contient des PIDs identifiant les propriétés d'intérêt dont les arêtes seront traversées. Ici, nous considérons `instance of` (P31), `subclass of` (P279) et `part of` (P361). L'indication du PID d'une propriété conduit à traverser des arêtes directes tandis que l'indication (-)PID conduit à traverser des arêtes inverses. Ici, les arêtes directes et inverses de P279 seront parcourues. L'écran d'entrée de KGPrune est présenté sur la figure 1.a.

qid_example.csv	pid_example.csv
Q18833	P31
Q251	P279
	(-)P279
	P361

Table 1. Exemple de fichiers d'entrée pour KGPrune. Le fichier `qid_example.csv` contient les QIDs des entités de départ d'intérêt. Leur voisinage sera récupéré en parcourant les arêtes étiquetées par les propriétés dont les PIDs sont spécifiés dans `pid_example.csv`.

Extraction de sous-graphes.

Une fois les fichiers CSV d'entrée soumis, KGPrune exécute notre algorithme de parcours et d'élagage [10][10]. À partir des entités de départ, les arêtes étiquetées par les propriétés spécifiées sont parcourues. Pour chaque voisin, notre modèle d'élagage analogique décide de le conserver ou de l'élaguer. Les analogies sont des énoncés de la forme A est à B ce que C est à D , modélisés comme des quadruplets $A : B :: C : D$ tels que $\text{Paris} : \text{France} :: \text{Berlin} : \text{Allemagne}$. De tels quadruplets capturent les similarités et les dissimilarités entre les objets [16, 17]. Ici, étant donné une entité de départ e_s^u spécifiée par l'utilisateur et un de ses voisins e_r^u, notre modèle prédit s'ils forment une analogie avec une entité de départ e_s^k et un de ses voisins e_r^k pour lesquels une décision de conservation est connue :

$$\underbrace{e_s^k : e_r^k}_{\text{Décision de « conservation » connue}} :: \underbrace{e_s^u : e_r^u}_{\text{Décision inconnue}} \tag{1}$$

Cette prédiction repose sur des plongements de graphe préappris pour les entités et sur le modèle convolutionnel de détection d'analogie introduit par [13]. Grâce à son architecture, ce modèle est capable de capturer les similarités et les dissimilarités relatives entre les entités de départ et leurs voisins à conserver ou à élaguer, et est ainsi capable de généraliser à des entités hétérogènes non-vues à l'entraînement. Si notre modèle prédit que les quatre entités mentionnées précédemment forment une analogie, la décision connue entre e_s^k et e_r^k (i.e. garder e_r^k) est extrapolée à e_r^u. Sinon, e_r^u est élaguée. Essentielles au modèle, les décisions connues proviennent d'un ensemble de données annoté manuellement nommé `dataset1` et publiquement accessible.[11]

10. https://github.com/Orange-OpenSource/analogical-pruning.
11. https://doi.org/10.5281/zenodo.8091584.

Ce processus est effectué de manière itérative sur le voisinage des nœuds conservés jusqu'à ce qu'aucun autre voisin ne puisse être atteint. Les résultats sont ensuite affichés à l'utilisateur (figure 1.b) qui peut choisir de visualiser les sous-graphes extraits (figure 1.c) ou de les télécharger au format JSON ou RDF pour les importer dans un nouveau GC. L'interface de visualisation permet aux utilisateurs d'explorer les voisinages des entités de départ et d'évaluer les résultats de l'élagage. En particulier, les utilisateurs peuvent remarquer dans l'interface utilisateur si notre modèle a élagué à tort un voisin qui les intéresse et l'ajouter aux entités de départ pour forcer sa prise en compte. Cela ouvre la voie à un processus d'élagage itératif dans lequel les utilisateurs explorent les résultats de l'élagage et ajoutent de nouvelles entités de départ qui seront exploitées dans les itérations suivantes.

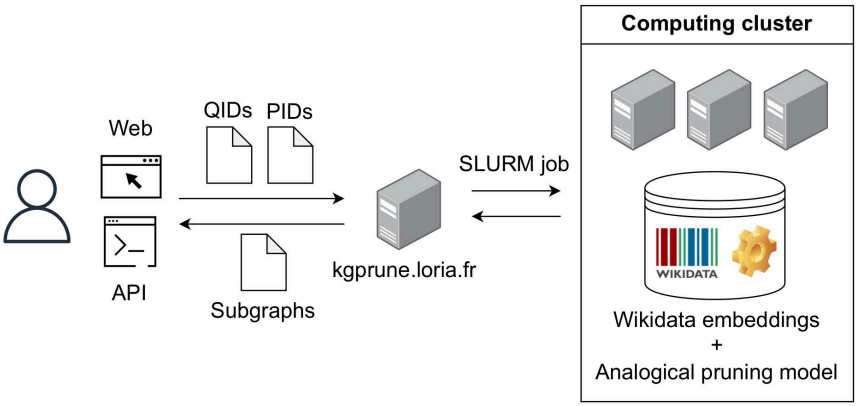


Fig. 2. Architecture technique de KGPrune. Les utilisateurs peuvent soumettre des tâches d'extraction de sous-graphes via le site Web ou l'API. Ces tâches sont ensuite soumises sous forme de tâches SLURM à notre cluster de calcul.

Architecture technique.

KGPrune repose sur l'architecture technique présentée sur la figure 2. Les utilisateurs peuvent interagir avec l'application via un navigateur Web ou une API. Leurs tâches d'extraction de sous-graphes sont envoyées sous forme de tâches SLURM à notre cluster de calcul où l'adjacence du graphe Wikidata, les plongements pré-entraînés, et nos modèles d'élagage analogique sont chargés et utilisés en inférence.

Pour l'apprentissage des plongements de Wikidata, nous avons utilisé le modèle TransE [2] avec une dimension de 200. Pour le modèle analogique, nous l'avons entraîné en utilisant `dataset1`, un des deux jeux de données que nous

avons préalablement annotés manuellement et rendus publiquement accessibles. Nous utilisons 16 filtres sur la première couche convolutionnelle et 8 filtres sur la deuxième couche convolutionnelle.

Notre modèle atteint des performances compétitives par rapport aux principaux modèles de l'état de l'art, avec un nombre de paramètres considérablement inférieur et une capacité de généralisation supérieure dans un contexte de transfert (table 2).

Modèle		**LSTM**	**Path Analogy**
dataset1	Précision	79.72 ± 5.17	80.10 ± 0.84
	Rappel	76.00 ± 6.59	74.44 ± 5.28
	F1	77.43 ± 2.38	77.06 ± 2.89
	ACC	83.48 ± 3.05	83.51 ± 2.87
	# paramètres	210,751	1,401
dataset2	Précision	78.49 ± 8.80	81.63 ± 8.27
	Rappel	94.58 ± 2.96	94.90 ± 2.16
	F1	85.36 ± 4.53	87.54 ± 5.05
	ACC	78.66 ± 5.95	82.50 ± 6.07
	# paramètres	210,751	251
Transfert	Précision	92.83	91.49
	Rappel	74.73	83.39
	F1	82.80	87.25
	ACC	80.04	84.33

Table 2. Performances de notre modèle comparé à un LSTM, son principal concurrent. Le contexte de transfert correspond à entraîner le modèle sur dataset1 et à le tester sur dataset2. Les expériences sur chaque jeu de données ont été réalisées en utilisant une validation croisée en 5 blocs et avec un hypertuning du nombre de filtres. Les résultats complets sont disponibles dans [10].

Deux cas d'usage illustratifs

Pour illustrer l'impact de l'application KGPrune, nous l'avons expérimentée sur deux cas d'utilisation : l'amorçage d'un GC d'entreprise et l'extraction de sous-graphes liés à des œuvres d'art pillées, attestant l'utilité de notre outil sur de réelles applications.

Amorçage d'un GC d'entreprise (GCE).

La construction d'un nouveau GC nécessite son amorçage avec un noyau de qualité qui peut ensuite soutenir des approches d'extraction automatique de connaissances à partir de données structurées ou non structurées (par exemple, tables, textes) [14, 20, 25]. Ces approches enrichissent ensuite le GC tout en étant guidées par les termes et les relations que le GC fournit, formant une boucle vertueuse.

Pour construire un tel noyau, plusieurs auteurs s'appuient sur Wikidata. Pour limiter la taille du noyau créé, des parties du voisinage des entités de départ d'intérêt sont sélectionnées avec un processus de distillation [21] ou d'élagage [10, 11]. Leurs parcours du graphe se concentrent sur la hiérarchie de l'ontologie, uniquement vers la racine [21] ou à la fois vers la racine et vers les feuilles [11].

Dans [10], nous avons proposé notre approche d'élagage pour amorcer un GCE centré sur le domaine de l'informatique, en parcourant l'ontologie vers la racine et vers les feuilles, à partir des entités de départ d'intérêt disponibles dans le glossaire interne de l'entreprise. Les performances obtenues par notre approche (table 2) sont très compétitives par rapport aux autres modèles d'état de l'art, avec un nombre de paramètres 100 à 1000 fois moindre. Cela illustre l'intérêt de notre approche pour ce cas d'utilisation. Avec KGPrune, nous avons étendu notre approche précédente en permettant à l'utilisateur de définir les propriétés à parcourir, enrichissant ainsi les sous-graphes extraits de Wikidata, et en offrant des capacités de visualisation pour permettre à l'utilisateur d'explorer les voisinages extraits.

Extraction de sous-graphes liés aux œuvres d'art pillées.

Les réseaux de pillage d'œuvres d'art opèrent à de nombreux niveaux et sur de longues périodes de temps. Certaines agences soulignent qu'il s'agit d'une industrie criminelle qui rapporte des milliards de dollars par an. Une documentation fiable est donc de la plus haute importance pour retrouver des biens culturels perdus ou volés et pour établir la propriété légitime. Il s'agit cependant d'une tâche difficile car les données sur le patrimoine culturel sont enfermées dans des silos de données, ce qui rend exceptionnellement difficiles la recherche, la localisation et l'obtention d'une documentation fiable [8].

Les auteurs de [8] proposent l'utilisation des données ouvertes et liées (en anglais, *Linked Open Data* ou LOD) comme base de données mondiale sur le patrimoine culturel. Ils ont exploré le potentiel des LOD pour intégrer de grandes quantités de données sur le patrimoine culturel, facilitant l'accès à l'information dans ce domaine, afin d'aider à protéger les biens culturels contre le pillage et de suivre les œuvres d'art pillées [26, 27]. Cependant, le suivi des œuvres d'art volées implique des connaissances relatives aux œuvres d'art, à la

généalogie, à la propriété, à la provenance. Certaines de ces connaissances sont présentes dans des GC génériques, mais associées à des connaissances non pertinentes pour la tâche considérée (e.g. biologie, informatique). Par exemple, Wikidata contient 43 730 marchands d'art, collectionneurs, conservateurs et galeries ; 6 648 musées d'art ; 930 405 peintures ; 251 personnes sur lesquelles l'Art Looting Investigation Unit[12] a enquêté ; et des propriétés telles que `propriétaire de` et `appartenant à`. D'où la nécessité d'extraire des sous-graphes spécifiques traitant de thèmes pertinents tout en évitant les informations fausses, inexactes ou non pertinentes.

Dans cette optique, KGPrune a le potentiel d'extraire et de collecter des informations fiables et pertinentes à partir de Wikidata. Dans des expériences préliminaires, nous avons appliqué notre approche sur le voisinage d'œuvres d'art connues (e.g. `Les Cyprès`), d'artistes (e.g. `Alexej von Jawlensky`), de musées (e.g., `National Gallery of Arts`) et de marchands d'art (e.g. `Alfred Flechtheim`). Les résultats ont montré une bonne adéquation avec les besoins humains en informations nécessaires au suivi des œuvres d'art volées. Nous sommes en train d'explorer plus en détail comment KGPrune, et en particulier ses fonctionnalités d'élagage et de visualisation, peuvent prendre en charge d'autres cas d'utilisation liés au patrimoine culturel.

Conclusion et perspectives

Dans cet article, nous avons présenté KGPrune, une application Web permettant aux utilisateurs d'extraire des sous-graphes d'intérêt de Wikidata en fournissant des entités de départ d'intérêt et des propriétés à parcourir. Notre application évite une potentielle dérive thématique lors du parcours du graphe en s'appuyant sur un mécanisme d'élagage efficace basé sur le raisonnement par analogie. Les utilisateurs peuvent interagir avec KGPrune via leur navigateur Web ou une API, facilitant ainsi son intégration dans des projets divers. Nous avons démontré l'intérêt de l'application avec deux cas d'utilisation concrets.

Actuellement, KGPrune ne prend en charge que Wikidata. Nous envisageons à l'avenir d'intégrer des GC supplémentaires (par exemple, DBpedia [12], YAGO [23], Bio2RDF [6]), fournissant de ce fait aux utilisateurs des contextes supplémentaires à partir desquels extraire des sous-graphes. De plus, notre modèle basé sur l'analogie est entraîné sur un jeux de données manuellement annoté et constitué d'entités de départ et de voisins à conserver ou à élaguer. Même si les expériences et les cas d'utilisation ont mis en évidence la capacité de généralisation de notre modèle, il est possible que la définition apprise des décisions de conservation ou d'élagage d'entités voisines ne soit pas adaptée à d'autres applications. Pour répondre à cette question, nous prévoyons de

12. `https://www.wikidata.org/wiki/Q30335959`.

permettre aux utilisateurs de fournir leurs propres exemples de voisins conservés et élagués. Ces exemples pourraient être utilisés dans la phase d'inférence ou même pour entraîner à la volée des modèles personnalisés, étant donné la complexité réduite de nos modèles. Pour guider ces développements futurs, nous sommes à l'écoute des retours d'utilisation, cas d'usage, contextes applicatifs, et thèmes d'intérêt de la communauté.

Remerciements

Ces travaux sont soutenus par le projet AT2TA[13] (ANR-22-CE23-0023) financé par l'Agence nationale de la recherche (ANR).

Références

[1] Tim Berners-Lee, James Hendler, and Ora Lassila. 2001. The semantic web. *Scientific american* 284, 5: 28–37.

[2] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems 26: 27th annual conference on neural information processing systems 2013. Proceedings of a meeting held december 5-8, 2013, lake tahoe, nevada, united states*, 2787–2795. Retrieved from https://proceedings.neurips.cc/paper/2013/hash/1cecc7a77928ca8133fa24680a88d2f9-Abstract.html.

[3] Jiaoyan Chen, Hang Dong, Janna Hastings, Ernesto Jiménez-Ruiz, Vanessa López, Pierre Monnin, Catia Pesquita, Petr Skoda, and Valentina A. M. Tamma. 2023. Knowledge graphs for the life sciences: Recent developments, challenges and opportunities. *Transactions on Graph Data and Knowledge* 1, 1: 5:1–5:33. https://doi.org/10.4230/TGDK.1.1.5.

[4] Claudia d'Amato, Louis Mahon, Pierre Monnin, and Giorgos Stamou. 2023. Machine learning and knowledge graphs: Existing gaps and future research challenges. *Transactions on Graph Data and Knowledge* 1, 1: 8:1–8:35. https://doi.org/10.4230/TGDK.1.1.8.

[5] Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. 2014. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *The 20th ACM SIGKDD international conference on knowledge discovery and data mining, KDD '14, new york, NY, USA - august 24 - 27, 2014*, 601–610. https://doi.org/10.1145/2623330.2623623.

[6] Michel Dumontier, Alison Callahan, Jose Cruz-Toledo, Peter Ansell, Vincent Emonet, François Belleau, and Arnaud Droit. 2014. Bio2RDF release 3: A larger, more connected network of linked data for the life sciences. In *Proceedings of the ISWC 2014 posters & demonstrations track a track within the 13th international semantic web conference, ISWC 2014, riva del garda, italy, october 21, 2014* (CEUR workshop proceedings), 401–404. Retrieved from https://ceur-ws.org/Vol-1272/paper_121.pdf.

[7] Mariano Fernández-López, Asuncion Gomez-Perez, and Natalia Juristo. 1997. METHONTOLOGY: From ontological art towards ontological engineering. *Engineering Workshop on Ontological Engineering (AAAI97)*.

13. https://at2ta.loria.fr/.

[8] Eleanor E. Fink, Pedro A. Szekely, and Craig A. Knoblock. 2014. How linked open data can help in locating stolen or looted cultural property. In *Digital heritage. Progress in cultural heritaage: Documentation, preservation, and protection - 5th international conference, EuroMed 2014, limassol, cyprus, november 3-8, 2014. proceedings* (Lecture notes in computer science), 228–237. https://doi.org/10.1007/978-3-319-13695-0_22.

[9] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann. 2021. *Knowledge graphs*. Morgan & Claypool Publishers. https://doi.org/10.2200/S01125ED1V01Y202109DSK022.

[10] Lucas Jarnac, Miguel Couceiro, and Pierre Monnin. 2023. Relevant entity selection: Knowledge graph bootstrapping via zero-shot analogical pruning. In *Proceedings of the 32nd ACM international conference on information and knowledge management, CIKM 2023, birmingham, united kingdom, october 21-25, 2023*, 934–944. https://doi.org/10.1145/3583780.3615030.

[11] Lucas Jarnac and Pierre Monnin. 2022. Wikidata to bootstrap an enterprise knowledge graph: How to stay on topic? In *Proceedings of the 3rd wikidata workshop 2022 co-located with the 21st international semantic web conference (ISWC2022), virtual event, hanghzou, china, october 2022* (CEUR workshop proceedings). Retrieved from https://ceur-ws.org/Vol-3262/paper16.pdf.

[12] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, and Christian Bizer. 2015. DBpedia - A large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web* 6, 2: 167–195. https://doi.org/10.3233/SW-140134.

[13] Suryani Lim, Henri Prade, and Gilles Richard. 2019. Solving word analogies: A machine learning perspective. In *Symbolic and quantitative approaches to reasoning with uncertainty, 15th european conference, ECSQARU 2019, belgrade, serbia, september 18-20, 2019, proceedings* (Lecture notes in computer science), 238–250. https://doi.org/10.1007/978-3-030-29765-7_20.

[14] Jixiong Liu, Yoan Chabot, Raphaël Troncy, Viet-Phi Huynh, Thomas Labbé, and Pierre Monnin. 2023. From tabular data to knowledge graphs: A survey of semantic table interpretation tasks and methods. *Journal of Web Semantics* 76: 100761. https://doi.org/10.1016/J.WEBSEM.2022.100761.

[15] Farzaneh Mahdisoltani, Joanna Biega, and Fabian M. Suchanek. 2015. YAGO3: A knowledge base from multilingual wikipedias. In *Seventh biennial conference on innovative data systems research, CIDR 2015, asilomar, CA, USA, january 4-7, 2015, online proceedings*. Retrieved from http://cidrdb.org/cidr2015/Papers/CIDR15_Paper1.pdf.

[16] Laurent Miclet, Sabri Bayoudh, and Arnaud Delhay. 2008. Analogical dissimilarity: Definition, algorithms and two experiments in machine learning. *Journal of Artificial Intelligence Research* 32: 793–824. https://doi.org/10.1613/jair.2519.

[17] Melanie Mitchell. 2021. Abstraction and analogy-making in artificial intelligence. *Annals of the New York Academy of Sciences* 1505, 1: 79–101.

[18] Natalya Fridman Noy, Yuqing Gao, Anshu Jain, Anant Narayanan, Alan Patterson, and Jamie Taylor. 2019. Industry-scale knowledge graphs: Lessons and challenges. *Communications of the ACM* 62, 8: 36–43. https://doi.org/10.1145/3331166.

[19] Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. 2024. Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering*: 1–20. https://doi.org/10.1109/tkde.2024.3352100.

[20] Juan Sequeda and Ora Lassila. 2021. *Designing and building enterprise knowledge graphs*. Morgan & Claypool Publishers. https://doi.org/10.2200/S01105ED1V01Y202105DSK020.

[21] Basel Shbita, Anna Lisa Gentile, Pengyuan Li, Chad DeLuca, and Guang-Jie Ren. 2023. Understanding customer requirements - an enterprise knowledge graph approach. In *The semantic web - 20th international conference, ESWC 2023, hersonissos, crete, greece, may 28 - june 1, 2023, proceedings* (Lecture notes in computer science), 625–643. https://doi.org/10.1007/978-3-031-33455-9_37.

[22] Bill Swartout, Ramesh Patil, Kevin Knight, and Tom Russ. 1996. Toward distributed use of large-scale ontologies. In *Proceedings of the tenth workshop on knowledge acquisition for knowledge-based systems*, 25.

[23] Thomas Pellissier Tanon, Gerhard Weikum, and Fabian M. Suchanek. 2020. YAGO 4: A reason-able knowledge base. In *The semantic web - 17th international conference, ESWC 2020, heraklion, crete, greece, may 31-june 4, 2020, proceedings* (Lecture notes in computer science), 583–596. https://doi.org/10.1007/978-3-030-49461-2_34.

[24] Denny Vrandecic and Markus Krötzsch. 2014. Wikidata: A free collaborative knowledgebase. *Communications of the ACM* 57, 10: 78–85. https://doi.org/10.1145/2629489.

[25] Gerhard Weikum, Xin Luna Dong, Simon Razniewski, and Fabian M. Suchanek. 2021. Machine knowledge: Creation and curation of comprehensive knowledge bases. *Foundations and Trends Databases* 10, 2-4: 108–490.

[26] Laurel Zuckerman. 2020. Linked data and holocaust era art markets: Gaps and dysfunctions in the knowledge supply chain. In *Proceedings of the international conference collect and connect: Archives and collections in a digital age, leiden, the netherlands, november 23-24, 2020* (CEUR workshop proceedings), 13–24. Retrieved from https://ceur-ws.org/Vol-2810/paper2.pdf.

[27] Laurel Zuckerman. Tracking looted art with graphs: A case study. Retrieved from https://api.semanticscholar.org/CorpusID:247314664.

Oublions Lena

Lena Forsén fait valoir ses droits à la retraite

Florence Sèdes[1]

L'« objectification », phénomène formalisé par les études en philosophie, est une théorie qui met l'accent sur les processus par lesquels les personnes sont traitées comme des objets, l'objet prenant le pas sur la personne dans son ensemble. Les analyses contemporaines, par exemple en sociologie, font que cette théorie est de nos jours très liée à l'objectification du corps des femmes. Cette objectification s'illustre dans le cadre de l'exposition visuelle, rendant cette réification évidente dans le cadre de la photographie (le fameux calendrier Pirelli) ou du cinéma (le *male gaze* où le personnage féminin n'est qu'un faire-valoir de son homologue, ce qui importe est ce que l'on voit d'elle ou ce que le personnage masculin pense d'elle, mais pas son ressenti personnel).

Lorsque dans les années 70, l'image de Lena Forsén est utilisée pour la première fois comme image de référence en traitement d'images, on est loin de cette prise de conscience. Cinquante ans plus tard, le message diffusé par l'IEEE[2] exige de remplacer l'image en question par une autre car elle enfreint le code de déontologie (adopté en juin 2020) de l'association professionnelle.

Cinquante ans d'exploitation de cette image prennent fin, et comme l'annonce Lena elle-même : « Once upon a time, I was the centerfold of Playboy. But I retired from modelling a long time ago. It's time I retire from tech, too » !

1. Professeure d'informatique à l'université Toulouse 3.
2. https://ieeefrance.org/a-propos-de-ieee/.

DOI:10.48556/SIF.1024.24.189

Il était une fois…

Recadrée à partir des épaules, la photo centrale de Playboy du mannequin suédois Lena Forsén regardant le photographe de dos fut l'étalon improbable des recherches de la communauté du traitement d'image, et l'une des images les plus reproduites de tous les temps. « Miss November », playmate d'un jour, aura vu son unique cliché pour le magazine sublimé… bien malgré elle.

Peu après son impression dans le numéro de novembre 1972 du magazine PlayBoy, la photographie a été numérisée par Alexander Sawchuk, professeur assistant à l'université de Californie, à l'aide d'un scanner conçu pour les agences de presse. Sawchuk et son équipe cherchaient de nouvelles données pour tester leurs algorithmes : la fameuse page centrale du magazine fut élue, et ce choix justifié par la présence d'un visage et d'un mélange de couleurs claires et foncées. Heureusement, les limites du scanner ont fait que seuls les cinq centimètres encadrant le visage ont été scannés, l'épaule nue de Forsén laissant deviner la nature de l'image originale, à une époque où la pornographie et la nudité étaient jugées à des niveaux différents de ceux où elles le sont désormais.

Madone en image processing…

Dès lors, la photo est devenue une image de référence standard, utilisée un nombre incalculable de fois depuis plus de 50 ans dans des articles pour démontrer les progrès de la technologie de compression d'images, tester de

nouveaux matériels et logiciels et expliquer les techniques de retouche d'images. L'image aurait même été une des premières à être déposée sur ARPANET, son modèle, Lena, ignorant tout de cette soudaine et durable célébrité.

Lena, véritable étudiante suédoise à New York, modèle d'un jour, a enfin le droit, en cette fin de l'année 2023, de faire valoir son droit à l'oubli : l'IEEE a publié un avis à l'intention de ses membres mettant en garde contre l'utilisation continue de l'image de Lena dans des articles scientifiques.

« *À partir du 1er avril, les nouveaux manuscrits soumis ne seront plus autorisés à inclure l'image de Lena* », a écrit le vice-président de l'IEEE Computer Society. Citant une motion adoptée par le conseil d'édition du groupe :

> « *La déclaration de l'IEEE sur la diversité et les politiques de soutien telles que le code d'éthique de l'IEEE témoignent de l'engagement de l'IEEE à promouvoir une culture inclusive et équitable qui accueille tout le monde. En accord avec cette culture et dans le respect des souhaits du sujet de l'image, Lena Forsén, l'IEEE n'acceptera plus les articles soumis qui incluent l'image de Lena* ».[3]

L'IEEE n'est pas la première à « bannir » cette photo de ses publications : en 2018, *Nature Nanotechnology* avait publié une déclaration interdisant l'image dans toutes ses revues de recherche, écrivant dans un éditorial que « … l'histoire de l'image de Lena va à l'encontre des efforts considérables déployés pour promouvoir les femmes qui entreprennent des études supérieures en sciences et en ingénierie… ».

De multiples raisons scientifiques ont été invoquées pour expliquer cette constance dans l'utilisation de l'image-étalon, rare dans nos domaines : la « gamme dynamique » (nombre de couleurs ou de niveaux de gris utilisées dans une image), la place centrale du visage, la finesse des détails des cheveux de Lena et la plume du chapeau.

En 1996, une note du rédacteur en chef de *IEEE Transactions on Image Processing* (à l'époque où cette photo était apparue) indiquait, pour expliquer pourquoi, selon lui, cette image avait été choisie et était devenue un « standard » que « l'image de Lena est celle d'une femme attirante », ajoutant : « Il n'est pas surprenant que la communauté des chercheurs en traitement d'images [essentiellement masculine] ait gravité autour d'une image qu'elle trouvait attrayante ».[4]

---

3. IEEE author center Article Submission Requirements — Article Submission Requirements […] Graphics Requirements […] Lena image: *IEEE's diversity statement and supporting policies such as the IEEE Code of Ethics speak to IEEE's commitment to promoting an inclusive and equitable culture that welcomes all. In alignment with this culture and with respect to the wishes of the subject of the image, Lena Forsén, IEEE will no longer accept submitted papers which include the "Lena image."*

4. *« […] the Lena image is a picture of an attractive woman […] It is not surprising that the (mostly male) image processing research community gravitated toward an image that they found attractive […] »*, David C. Munson, A note on Lena, IEEE trans… proc, 5/1 1996-01.

Le traitement de l'image qui ignore la « sémantique » de l'objet central, objectifiant l'individu qui pose.

Le magazine PlayBoy aurait pu lui-même mettre un terme à la diffusion de l'image de Lena : en 1992, le magazine avait menacé d'agir, mais n'a jamais donné suite. Quelques années plus tard, la société a changé d'avis : « nous avons décidé d'exploiter ce phénomène », a déclaré le vice-président des nouveaux médias de Playboy en 1997.

Lena Forsén elle-même, « sainte patronne des JPEG » a également suggéré que la photo soit retirée. Le documentaire *Losing Lena* a été le déclencheur pour encourager les chercheurs en informatique à passer à autre chose[5].

Fabio is the new Lena

Fabio Lanzoni, top model italien, sera, le temps d'une publication, le « Lena masculin » : dans l'article « Stable image reconstruction using total variation minimization », publié en 2013, Deanna Needell et Rachel Ward décident d'inverser la vision du « gender gap » (inégalités de genre) en choisissant un modèle masculin.

La légende a débordé du cadre purement académique : en 2016, « Search by Image, Live (Lena/Fabio) », de l'artiste berlinois Sebastian Schmieg, utilise le moteur de recherche inversée d'images de Google pour décortiquer les récits de plus en plus nombreux autour de l'image célèbre de Lena[6] : l'installation est basée sur une requête lancée avec l'image de Lena vs. une requête duale lancée avec l'effigie du blond mâle Fabio. Son objectif est d'analyser la manière dont les technologies en réseau façonnent les réalités en ligne et hors ligne. Beau cas d'usage pour la « story » de notre couple !

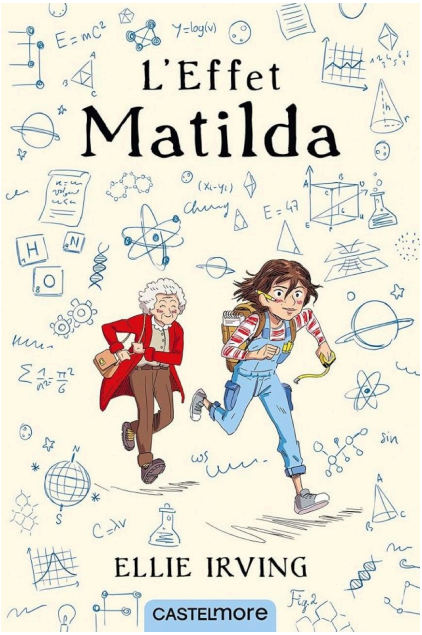
5. https://vimeo.com/372265771.

6. https://thephotographersgallery.org.uk/whats-on/sebastian-schmieg-search-image-live-lena-fabio.

Lena vs. Matilda ?

Alors qu'on parle d'effet Matilda et d'invisibilisation des scientifiques, pour le coup, voilà une femme très visible dans une communauté où les femmes sont sous-représentées !

Quel message renvoie cet usage d'une photo « légère », indéniablement objectifiée, pour former des générations d'informaciens ? Comment expliquer l'usage par une communauté d'un matériel déshumanisé dont on oublie symbolique et sémantique, alors que le sujet pourrait être considéré comme dégradant pour les femmes ?



Comment interpréter l'usage abusif, irrespectueux du droit d'auteur, du consentement et de l'éthique, par une communauté très masculinisée d'une seule et unique image féminine ? Effet de halo, biais de confirmation ou de représentativité ? L'ancrage du stéréotype est ici exemplaire, avec cet amplificateur qu'ont constitué l'usage d'une communauté et la généralisation des compétitions d'algorithmes, des évaluations de jeux de données, et la propagation de « l'objet » Lena.

L'évolution de l'objectification, au travers de cette spectaculaire illustration du *male gaze*, et de l'utilisation pendant des décennies d'un objet iconique dont peu se sont offusqués, adresse la transversalité des sciences[7] et la nécessité de réfléchir ensemble à la conception, la manipulation et l'agrégation des données, réelles ou synthétiques.

Alors, conformément aux préconisations de l'IEEE, remercions Lena d'avoir permis les progrès des algorithmes de traitement d'images. Engageons-nous désormais à l'oublier, marquant ainsi « un changement durable pour demain », et à encourager, inspirer, accueillir, connecter toutes les futures générations de femmes scientifiques.

7. https://webcms.i3s.unice.fr/TRACTIVE/.

Le 60e anniversaire du langage Basic

Pierre Mounier-Kuhn[1]

Le langage Basic a été, à l'époque où l'informatique personnelle émergeait – fin des années 70 et début des années 80 –, LE langage de programmation « fourni » avec tous les micro-ordinateurs, souvent le seul langage disponible, d'ailleurs ; mais sait-on qu'en cette année 2024, on fête le soixantième anniversaire de Basic ?…

Au début des années 1960, la diffusion des ordinateurs dans les universités américaines et européennes incitait un nombre croissant d'étudiants à apprendre à programmer. La plupart voulaient, non pas devenir informaticiens, mais seulement pouvoir utiliser ces machines pour effectuer des calculs ou, plus tard, des tâches de gestion dans leur vie professionnelle.

On ne pouvait utiliser pour cela ni le langage machine formé de chaînes binaires, ni les codes « assembleur » qui restaient plutôt ésotériques et spécifiques à chaque modèle d'ordinateur.

Plusieurs langages spécialisés de haut niveau avaient déjà été inventés : Fortran pour le calcul scientifique, Cobol pour la gestion comptable, Lisp pour l'intelligence artificielle, etc. Le désir d'un langage universel, à la fois rigoureux et assurant la portabilité des programmes quel que soit le modèle d'ordinateur, motivait le développement d'Algol par un collectif de scientifiques de différentes disciplines — mathématique, astronomie, informatique —, produisant à la fois un outil de programmation efficace et le premier programme de recherche en informatique. Conçus pour les usages professionnels, tous ces systèmes de codage étaient peu adaptés à l'initiation à l'informatique.

1. Historien, chargé de recherche au CNRS.

DOI:10.48556/SIF.1024.24.195

L'invention du langage Basic au début des années soixante par des mathématiciens du Dartmouth College est venue combler ce manque.

Dartmouth College est un prestigieux établissement privé d'enseignement secondaire et supérieur du New Hampshire. À l'époque, deux mathématiciens, John Kemeny et Thomas Kurtz, ont entrepris d'y réaliser un système d'exploitation en temps partagé (*time-sharing*), permettant de connecter plusieurs terminaux à l'ordinateur du centre de calcul. C'est le meilleur moyen d'enseigner la programmation à un groupe d'élèves : chacun disposera de son terminal et aura directement accès à l'unité centrale pour tester son programme. C'est une forme d'informatique personnelle, dix ans avant les micro-ordinateurs.

Sous la direction de Kemeny et Kurtz, au premier semestre 1964, une équipe d'étudiants développe à la fois le système d'exploitation et un nouveau langage : *Beginner's All-purpose Symbolic Instruction Code* (Basic). Beaucoup plus simple que Fortran, Basic est à la fois plus rapide à traiter par la machine et plus facile à assimiler par des apprentis. Il a aussi un champ d'application plus large, pouvant servir aussi bien à des non-scientifiques : parmi les premiers utilisateurs figurent des étudiants du laboratoire de psychologie de Dartmouth. Le langage doit permettre l'émission de messages d'erreur compréhensibles, pour aider le novice à corriger son programme sans avoir besoin de consulter le manuel. Il recèle aussi des possibilités d'extension offrant des aptitudes de programmation plus avancées.

Fig. 1. John Kemeny lors d'un cours de programmation Basic à l'université de Dartmouth. – Crédit : Dartmouth University.

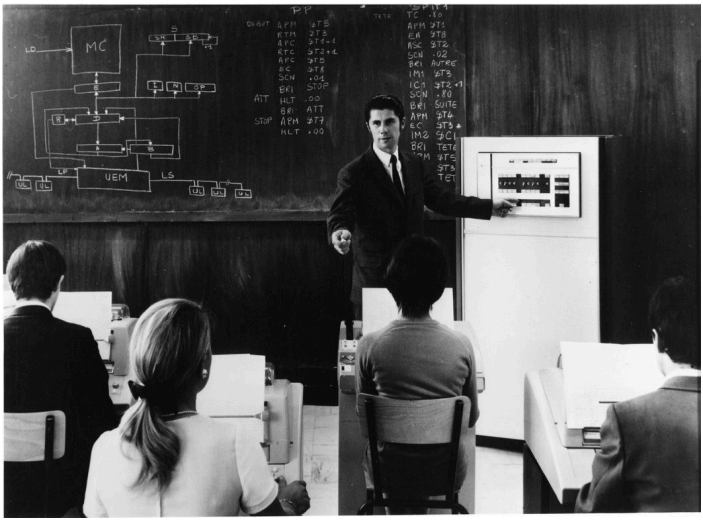


Fig. 2. Cours de programmation en France vers 1970 : les étudiants travaillent sur des terminaux « Teletype », sans écran (source : CII).

En quelques mois, le langage est défini et l'équipe rédige son compilateur. Basic repose sur une syntaxe minimaliste de quinze instructions (moitié moins que Fortran), comme :

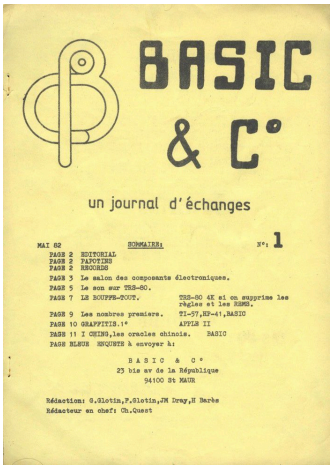
– **goto** : dérouter l'exécution vers une certaine ligne du programme (oui, Basic était un peu *line-oriented*, avec des lignes numérotées) ;
– **if… then…** : branchement conditionnel, dérouter conditionnellement le contrôle du programme vers une certaine ligne ;
– **print** : imprimer (la fonction « afficher à l'écran » n'aura de sens que dix ans plus tard, quand on commencera à connecter des écrans aux ordinateurs) ;
– **end** : fin du programme.

À partir de la rentrée 1964, la plupart des étudiants de Dartmouth s'initient à la programmation en Basic, ainsi qu'à l'emploi du système d'exploitation via les terminaux. Deux heures de cours et quelques travaux pratiques y suffisent. Ils peuvent effectuer des calculs, utiliser des logiciels d'application stockés dans une programmathèque sur bandes magnétiques — par exemple des logiciels éducatifs, des simulations, voire des jeux.

À la fin des années soixante, Basic est enseigné comme langage d'initiation dans des dizaines d'établissements américains. Il n'est pas le seul de son espèce : dès 1961 en France, la « Programmation automatique des formules »

(PAF), comparable à Basic par sa conception simple et son interactivité, est diffusée avec le mini-ordinateur CAB 500 de la Société d'électronique et d'automatisme ; ensuite apparaissent des langages plus élaborés comme Pascal, dérivé d'Algol, ou LSE (Langage symbolique d'enseignement) inventé à l'École supérieure d'électricité. Ces langages, notamment LSE, seront utilisés dans l'enseignement de l'informatique dans les lycées dès les années 1970.

Mais, de tous ces langages, Basic est celui qui connaît la plus grande diffusion. Tout constructeur d'ordinateurs voulant vendre son matériel sur le marché éducatif doit fournir un interpréteur Basic. D'abord voué à l'enseignement, ce langage rencontrera un succès important dans le monde des hobbyistes en raison de sa simplicité. En 1975, il deviendra le principal langage utilisable sur les micro-ordinateurs, bientôt associé aux produits d'un jeune éditeur de logiciels : Microsoft.


BASIC & C°

un journal d'échanges

MAI 82 SOMMAIRE: N°: 1

PAGE 2 EDITORIAL
PAGE 2 PANOTINS
PAGE 2 RECORDS
PAGE 3 Le salon des composants électroniques.
PAGE 5 Le son sur TRS-80.
PAGE 7 LE SOUFFE-TOUT. TRS-80 4K si on supprime les règles et les REMS.
PAGE 9 Les nombres premiers. TI-57,HP-41,BASIC
PAGE 10 GRAFFITIS.!* APPLE II
PAGE 11 I CHING,les oracles chinois. BASIC
PAGE BLEUE REQUETE à envoyer à:

B A S I C & C °
23 bis av de la République
94100 St MAUR

Rédaction: G.Slotin,F.Slotin,JM Drwg,H Barbe
Rédacteur en chef: Ch.Quest

Fig. 3. *Basic and C°*, n°1 (1982) et *Applications familiales en Basic*.

Basic a pu être critiqué pour son manque de rigueur, notamment pour l'abus de l'instruction « **goto**… » Mieux vaut ne pas utiliser ce langage pour programmer les commandes de vol d'un avion ou d'une centrale nucléaire ! Mais il a joué un rôle culturel décisif, en permettant à des milliers, puis à des millions d'utilisateurs, d'écrire leurs propres programmes et d'interagir avec un ordinateur.

Un accessoire des ordinateurs — toujours indispensable à l'heure actuelle sous l'un de ses nombreux avatars — est lui aussi apparu au grand public au début des années 1980 : la souris. Elle a été inventée par Douglas Engelbart exactement la même année que Basic, en 1964, à l'université de Standford à l'autre bout des États-Unis. Mais ceci est une autre histoire…

L'intelligence artificielle n'est pas une question technologique

Bruno Mermet[1]

Laurent BIBARD & Nicolas SABOURET
L'intelligence artificielle n'est pas une question technologique
Échanges entre le philosophe et l'informaticien
Éditions « l'aube »
ISBN 978-2-8159-5422-8
Date de parution : 21 avril 2023
144 pages, 115 mm × 215 mm, 17 €

Après une longue traversée du désert, depuis quelques années, avec l'avènement du *deep learning*, l'intelligence artificielle reprenait du poil de la bête. Toutefois, les avancées restaient plutôt cantonnées à la recherche et au monde socio-économique. Mais en 2022, deux logiciels d'IA générative sont devenus accessibles au grand public : au printemps, c'est Midjourney, un logiciel de génération d'image qui devient disponible pour tout le monde ; puis à l'automne, c'est au tour de ChatGPT, qu'on ne présente plus.
Ces deux logiciels ont séduit et conquis le grand public dans leurs domaines respectifs à une vitesse encore jamais vue, et les discussions sur l'intelligence artificielle envahissent la place publique, tout le monde se proclamant **expert** du domaine. Mais force est de constater que les prétendus experts de la technologie

1. Université Le Havre-Normandie.

CC BY NC SA
DOI:10.48556/SIF.1024.24.199

mise en œuvre ne le sont en général pas, et les débats sociétaux manquent souvent d'éléments scientifiques sur le sujet : c'est typiquement ce que l'on peut observer dans certaines émissions de télévision où il n'est pas rare que l'on réduise l'IA à l'IA générative.

Le livre *« L'intelligence artificielle n'est pas une question technologique »,* de Laurent Bibard et Nicolas Sabouret, vient à point pour enrichir le débat : il aborde enfin les enjeux du développement de l'IA en prenant en compte ses fondements scientifiques et ses aspects technologiques.

Contrairement à ce à quoi l'on pourrait s'attendre, le livre de Bibard & Sabouret ne se présente pas sous la forme d'un entretien, mais sous la forme d'une synthèse des réflexions des deux auteurs.

Le livre revient dans un premier temps sur les causes de l'aversion de beaucoup de gens vis-à-vis des outils informatiques : des entreprises qui remplacent des conseillers par des outils ne répondant qu'aux cas standard, car ces outils ont été mis en place dans le seul souci de réduire les coûts. Cela permet également de mettre en évidence la thèse développée par la suite que l'IA ne crée pas, et que la suppression de tout humain dans les systèmes de décision est problématique : non seulement une présence humaine permet de trouver des solutions à des problèmes nouveaux, mais elle peut aussi tre un garde-fou face à des décisions stupides, voire dangereuses, comme l'illustre la controverse Perrow-Weick que décrivent les auteurs.

Un passage nécessaire est consacré à la notion d'intelligence et au positionnement de l'intelligence artificielle par rapport à l'intelligence humaine. Là encore, des éléments clairs permettent de mieux comprendre en quoi l'IA diffère de l'intelligence humaine.

Les limites de l'IA sont également justifiées sur la base des résultats scientifiques sur la complexité (en temps et en espace) et l'indécidabilité, illustrés sur des exemples concrets et parlants. La démystification de l'IA est aussi renforcée en démontrant pourquoi cela n'a pas de sens de parler de **responsabilité** d'une IA.

Plusieurs aspects sociétaux sont également abordés. C'est par exemple le cas de l'influence de l'IA sur les emplois : l'ouvrage rappelle, à juste titre, que le but de tout développement technologique, depuis la création des premiers outils, à toujours été de simplifier la vie humaine, en facilitant, voire en supprimant, des tâches fastidieuses. L'IA n'est donc pas un cas particulier. L'addiction aux nouvelles technologies est également replacée dans le cadre général de l'effet de mode, niveau 0 de la sociabilisation, permettant ainsi de mieux remettre à sa place l'engouement constaté.

Cependant, l'ouvrage montre également que la méconnaissance du fonctionnement de ces outils peut amener à des mésusages risqués, et insiste sur la

nécessité, à tous les niveaux, de former à l'IA. Cette éducation doit être avant tout politique, mais également se retrouver tout au long du parcours éducatif des élèves.

Mon ressenti

Bien sûr, pour un informaticien un peu ouvert, l'ouvrage n'apporte pas vraiment de nouvelles connaissances ; ce n'est d'ailleurs pas son but. Mais il donne des clés pour mieux expliquer à des profanes tout ce qui se cache derrière ces fameuses IA. Et puis, surtout, l'ouvrage amène à se poser de nombreuses questions, et à revenir sur ses certitudes : la place de l'homme dans la nature, l'importance des émotions dans les prises de décision, l'autonomie…

Le retour sur le mythe de la caverne permet également de mieux comprendre les difficultés qu'il peut y avoir à sortir quelqu'un (un complotiste ?) de son erreur, et pourquoi l'éducation préalable est fondamentale.

L'interview

Un ouvrage qui comble un manque, une forme inattendue, un des auteurs, Nicolas Sabouret, collègue de la même communauté scientifique que moi… je voulais en savoir plus.

J'ai donc demandé aux deux auteurs de répondre **séparément** à mes questions. Mais — à l'instar de tout enseignant, qui exige de ses étudiants qu'ils respectent ses consignes — mes deux collègues ont préféré répondre **ensemble** ☺.

Bruno Mermet (BM) *Pourriez-vous vous présenter en quelques mots ?*

Laurent Bibard (LB) Je suis professeur à l'ESSEC, département Management, avec un cursus en philosophie et en gestion. Je travaille sur des questions de leadership, d'éthique et de complexité.

Nicolas Sabouret (NS) Je suis professeur à l'université Paris-Saclay en informatique et je travaille dans le domaine de la simulation des comportements humains pour la simulation multi-agents et pour les agents conversationnels animés.

BM *Comment vous êtes-vous rencontrés ? Comment est née l'idée de cet ouvrage ?*

NS En 2019, j'ai écrit un livre de vulgarisation qui s'appelle « Comprendre l'intelligence artificielle » et, à cette occasion, j'ai été invité à participer à un événement de vulgarisation scientifique à l'ESSEC. Il y avait un philosophe qui faisait un exposé que j'ai trouvé extrêmement intéressant sur ce qui caractérise l'être humain, avec une réflexion vis-à-vis des projections qu'on peut avoir sur l'intelligence artificielle et les machines. Et j'ai été conquis par la simplicité avec laquelle il présentait des notions que je n'arrivais pas bien à formuler et qui souvent me gênaient quand je devais répondre à des gens qui confondent intelligence humaine et intelligence artificielle. C'est ainsi qu'a commencé notre aventure commune.

LB La rencontre a été extrêmement agréable et vivante. Je trouve que Nicolas Sabouret est d'une clarté et d'une pondération exceptionnelle par rapport aux thèmes sur lesquels il est expert. Ainsi, le souhait est venu de collaborer : nos échanges me semblaient assez vivants et importants pour mériter d'être couchés sur le papier et de se transformer en livre.
NS Je me souviens qu'on prenait une fois tous les mois ou tous les deux mois un café ensemble pour échanger sur nos idées sur l'IA et la philo. Jusqu'au moment où tu as dit : « Mais pourquoi, on n'enregistre pas ce qu'on raconte ? »
LB Oui, et nous avions fait plusieurs interventions communes dans des colloques et autres.

BM *Plutôt que présenter vos réflexions sous la forme d'une discussion, d'une confrontation de points de vue, votre livre se présente sous la forme d'une synthèse de vos idées. Ce travail a-t-il été facile à réaliser ?*

LB Je crois qu'on peut dire que nous avions l'intention de garder la forme d'un dialogue, mais que ce sont les éditeurs, qui nous ont conseillé de ne pas le faire. Et c'est devenu une réflexion commune. Ce que je trouve très bien dans nos discussions, c'est qu'on peut avoir parfois des questions ou des désaccords que nous arrivons à creuser suffisamment pour que je comprenne des contenus techniques auxquels je ne connaissais rien et pour que nous partagions véritablement les enjeux de la réflexion.
NS Je me souviens d'un point sur lequel nous avons dû pas mal discuter : il s'agit de l'usage du terme « intelligence artificielle ». Je crois que tu ne voulais pas du tout qu'on parle d'intelligence artificielle et je pense que je t'ai convaincu en te donnant les raisons historiques de ce terme, qui est ancré dans une histoire et qui fait sens dans cette histoire (mais pas forcément dans le sens avec lequel il est parfois employé aujourd'hui en confondant technologie numérique et intelligence artificielle).
LB Certes, je comprends l'historique mais en revanche, quand j'interviens en conférence, je passe mon temps à dire qu'au sens fort, il n'y a pas d'intelligence artificielle, qu'il y a du calcul, qu'il est extraordinaire et que nous, humains, faisons une confusion.
NS D'ailleurs, quand j'interviens, je dis souvent : il n'y a pas d'intelligence dans l'intelligence artificielle, il n'y a pas d'apprentissage dans l'apprentissage automatique, et il n'y a pas de neurone dans les réseaux de neurones.

BM *Dans le cadre de la rédaction du livre, vous avez certainement appris beaucoup de vos échanges. Quel est, pour chacun d'entre vous, le principal enseignement que vous avez tiré de votre interaction avec l'autre ?*

LB Je dirais volontiers qu'il y en a tout le temps. Le fait que nous arrivions à dialoguer alors que moi, je n'y connaissais rien (j'insiste bien là-dessus) mais que

nous parvenions à tisser une compréhension partagée des enjeux que véhiculent les technologies autour de ce qu'on appelle l'IA, ça me paraît le point le plus saillant.

NS Je suis d'accord sur le fait que c'était un processus vraiment continu : à chaque fois qu'on s'est rencontrés et qu'on a discuté, nos échanges ont été un apprentissage réciproque. Et cela ne se voit pas dans le livre mais nous passions notre temps à se dire : « Ah oui, ce que tu dis, ça me parle, ça m'interpelle ». Nous étions constamment dans cette création de connaissances communes.

LB C'est un peu comme dans les négociations, quand elles sont bien menées. Tant qu'on reste dans l'opposition et à croire qu'on sait à l'avance ce qui va se passer, la négociation est plombée. Si, en revanche, on sait le mieux possible la position de chacun, ses attentes, son contexte, etc. et qu'en même temps on se dit : « je ne sais pas, et je vais écouter », alors émergent des espaces communs possibles pour dépasser la situation de conflit. Par analogie, je crois qu'on a été assez intéressés chacun par l'expertise, les compétences ou le regard de l'autre pour chaque fois, libérer un espace nouveau de réflexion qu'on n'attendait pas.

NS Je suis tout à fait d'accord avec toi et c'est d'ailleurs l'un des sujet de ton dernier bouquin, donc c'est pas surprenant que tu en parles ici !

LB Il y a un point saillant quand même dans ce que j'ai appris. j'adore la clarté avec laquelle tu dis : « il y a plus de parties d'échecs possibles que de nombre d'atomes dans l'univers visible » pour expliquer que l'IA avance par approximation et non pas de manière parfaite et exhaustive.

NS Moi ce que j'ai trouvé particulièrement intéressant (mais je m'y attendais car c'était la raison pour laquelle j'étais venu le trouver au départ), c'est ta réflexion sur le fait qu'il y a des catastrophes inévitables mais que la faillibilité de l'humain va lui permettre de rattraper ces catastrophes. Cette opposition entre la faiblesse de l'humain et sa capacité d'adaptation, je trouve que c'est un point crucial de la compréhension de ce qui nous différencie de l'intelligence artificielle.

BM *Pendant longtemps, philosophie et science étaient indissociables : les philosophes grecs étaient souvent des scientifiques, la « thèse » dans le monde anglo-saxon se nomme « Philosophiæ doctor » (PhD). En France, de nos jours, la philosophie est plus souvent vue comme une discipline littéraire, comme semble l'illustrer la spécialité du lycée « Humanités, littérature et philosophie ». Comment voyez-vous le lien entre science et philosophie ?*

LB Nous ne sommes pas du tout dans une approche scientifique de la philosophie. Nous sommes dans une approche humaine, qu'on peut du coup qualifier de philosophique, des sciences. Le thème que nous nous sommes donné est donc de comprendre les technologies à l'aide de la philosophie. Ce qui est vrai, c'est que la philosophie a été scientifique jusqu'à la Renaissance et qu'à partir de la Renaissance et l'explosion de l'usage des mathématiques (qui passe par

l'utilisation systématique de l'algèbre comme nous l'avons évoqué dans le livre), il y a une séparation des sciences et la philosophie est de plus en plus isolée. Un des effets est que la philosophie a mauvaise conscience, comme les sciences humaines : on aimerait arriver à une mathématisation pour avoir l'air aussi rigoureux. C'est à mon sens une énorme erreur parce que la philosophie est adossée évidemment à une compréhension des mathématiques mais à partir d'un usage préalable du langage. Nous en avons parlé un peu dans le livre : c'est la notion de « *logos* » en grec qui renvoie au langage et donc à la signification du monde qui est capitale dans l'exercice de la pensée. Au contraire le « ratio », la rationalité, réfère aux mathématiques, à la calculabilité des choses et ce n'est pas le même registre, mais ils se complètent. La philosophie à notre époque a une espèce de mauvaise conscience de croire qu'elle n'a plus de pertinence parce qu'elle n'est pas scientifique ; c'est une erreur de fond dans la compréhension de la philosophie.

NS Je suis tout à fait d'accord que notre livre est plutôt de la philosophie des sciences que des sciences de la philosophie. J'ajouterai que j'ai particulièrement apprécié la démarche qui consiste à comprendre la science pour pouvoir en discuter d'un point de vue philosophique. Je pense que nous partageons cette envie de comprendre les choses pour pouvoir en discuter : nous avons fait un exercice pédagogique réciproque pour pouvoir discuter sur des choses que nous avions comprises et que nous avions expliquées.

LB Et dont les enjeux de société sont essentiels ! Pour ma part, j'aime énormément comment nous avons écarté immédiatement le débat très mal posé entre technophilie et technophobie. Les deux sont absurdes. Les technologies par elles mêmes ne vont rien résoudre et les technologies par elles mêmes ne menacent rien : c'est strictement ce que nous en rêvons et ce que nous faisons, ce que nous en faisons, qui est en jeu. Et c'est pour ça qu'on a absolument besoin d'échanges entre philosophes et scientifiques. C'est absolument capital, parce que les enjeux sont politiques au sens noble du terme.

NS Je pense que, par ailleurs, nous avons tous les deux une culture pluridisciplinaire. Laurent entre philosophie et sciences de la gestion. Et moi entre informatique et sciences humaines (psychologie).

BM *Dans votre ouvrage, vous expliquez que l'intelligence humaine ne peut être reproduite par une machine. Quel est alors pour vous la faille dans le raisonnement suivant :*

- *l'intelligence humaine relève du fonctionnement du cerveau ;*
- *le cerveau est défini par un ensemble de neurones connectés entre eux et dont le fonctionnement est défini par des processus électro-chimiques ;*
- *tout processus électro-chimique peut être reproduit par un ordinateur.*

Donc l'intelligence humaine peut être reproduite par un ordinateur.

NS Pour moi la faille est à deux endroits. Premièrement, lorsqu'on dit « l'intelligence humaine relève du fonctionnement du cerveau », je pense que c'est plus compliqué que cela : l'intelligence humaine ne se limite pas à la rationalité cérébrale. Nos émotions et tout notre corps participent à notre intelligence, donc je pense qu'on ne peut pas forcément limiter au cerveau. Mais si on résume le cerveau au centre du système nerveux (et donc effectivement où toutes les émotions atterrissent), acceptons le point numéro un. Le principal problème pour moi est le troisième point : tout processus électrochimique ne peut pas être reproduit par un ordinateur, à moins que nous disposions d'un temps de calcul astronomique (choses dont nous ne disposons pas) ! Je pense que c'est là où il y a une faille : on s'imagine pouvoir tout simuler avec un ordinateur, en oubliant que chaque opération de simulation nécessite un certain temps de calcul.
LB Je crois que c'est absolument capital de l'expliciter très clairement pour qu'il n'y ait pas de fantasme. Mais de mon point de vue, la faille est sur le premier point. Je voudrais raconter l'anecdote suivante. Un jour, en formation permanente, une participante commence sa phrase en disant « Mon cerveau… ». Je l'arrête immédiatement de manière bienveillante, provocatrice et en plaisantant en disant : « Attendez, s'il y avait devant moi 30 cerveaux posés sur les tables, je pars en courant ! » Je ne peux pas imaginer enseigner à des cerveaux, j'enseigne à des gens qui sont là, concrets et on n'a jamais vu jusqu'à ce nouvel ordre biotechnologique un cerveau humain fonctionner en dehors d'un corps. Ainsi, nous pensons avec tout l'être que nous sommes et nous savons maintenant de mieux en mieux que nous avons des neurones partout, jusque dans l'estomac, chose que les taoïstes pressentaient depuis très longtemps : pour les taoïstes, manger et penser c'est exactement le même processus. Quand on dit : l'intelligence humaine relève du fonctionnement du cerveau, c'est une focalisation complètement obsessionnelle qui est très mal placée, très déséquilibrée, sur le cerveau comme organe exclusif. C'est donc là que se situe la faille principale à mon sens.

BM *Dans votre livre, vous opposez intelligence artificielle à intelligence humaine, sans évoquer plus généralement l'intelligence animale. Y a-t-il une raison à cela ? Où la placez-vous ?*

NS Je pense qu'il n'y avait pas de but de différencier l'intelligence humaine de l'intelligence animale. Je vois une continuité et nous parlions d'intelligence au sens général de l'être vivant. De plus, je pense qu'il est très difficile de définir l'intelligence humaine, mais il est encore plus difficile de définir l'intelligence animale : c'est aussi pour cela qu'il ne me semble pas judicieux que nous allions sur ce terrain-là (même si, une fois encore, il n'y avait pas de volonté de faire de distinction).
LB Je suis vraiment d'accord : ce n'était tout simplement pas le thème du bouquin. Nous aurions pu en parler, pourquoi pas, mais là ce n'était pas le

thème du livre. Ce n'est pas du tout une exclusion délibérée, c'est simplement qu'on traite d'une chose à la fois. S'il fallait par ailleurs évoquer la différence entre animalité et humanité, on pourrait dire que précisément, nous, humains, sommes des animaux dégénérés, en quelque sorte comme « sortis de la nature » partiellement (pas totalement, nous ne serions alors plus vivants) et du coup nous sommes fragiles, vulnérables à l'origine et donc aussi beaucoup plus adaptables, ne serait-ce que par nécessité, que les animaux. Ils sont dotés d'une mémoire biologique extrêmement importante par rapport à nous. Ils sont tellement plus adaptés que nous, comme préfabriqués pour un environnement donné, que si l'on détruit leur environnement, ils disparaissent (et c'est ce qu'on voit dramatiquement se passer actuellement). Tandis que nous sommes infiniment adaptables, y compris au pire (au regret de beaucoup d'historiens).

BM *J'ai posé la question suivante à chatGPT : « Laurent Bibard est professeur de philosophie et de gestion. Il est également docteur en philosophie. Nicolas Sabouret est professeur d'informatique, il mène des recherches sur les systèmes multi-agents. Ils décident d'écrire ensemble un livre intitulé « L'intelligence artificielle n'est pas une question technologique ». Propose-moi un résumé de ce livre ». Voici un résumé de la réponse que j'ai obtenue : « Le livre aborde des questions cruciales telles que : la définition de l'intelligence, l'impact sur le travail et l'économie, l'éthique et la responsabilité , la gouvernance de l'IA ». Votre avis ?*

NS Le résumé proposé ne correspond pas du tout au plan que nous avons adopté. C'est une autre preuve de ce que l'humain sort du cadre, là où la machine reste enfermée ou, comme dirait François Julien, « coïncidée » dans le passé. C'est pour ça qu'il faut une intelligence humaine pour sortir du cadre.

LB Il y a quelques mots qui, statistiquement peut-être, ressortent par le calcul et dont on pourrait dire « cela fait sens ». Mais en fait, la formulation ne fait pas sens du tout. Ce n'est pas ça que nous avons abordé.

NS Nous n'avons pas défini l'intelligence. Nous n'avons pas vraiment parlé de l'impact sur le travail et l'économie (on l'a évoqué dans un chapitre, mais c'est assez marginal). Nous n'avons pas parlé de l'éthique et de la responsabilité de l'IA, mais bien de l'éthique et de la responsabilité humaine, pas des machines. Enfin, la gouvernance de l'IA est très peu abordée, et certainement pas sous l'angle de « quelles sont les règles ». Donc ChatGPT n'aura pas une très bonne note ici !

BM *Merci beaucoup à tous les deux pour cet échange très instructif.*

Une bataille sans bataille

Jean-Paul Delahaye[1]

La rubrique « Récréation informatique » propose une petite énigme algorithmique ou sur un thème de mathématiques discrètes susceptible d'intéresser un lecteur de 1024. La solution est donnée dans le numéro suivant.

Rappel du problème précédent : *Vide ou pas à l'infini ?*

On dispose d'une urne vide et d'une réserve illimitée de billes. Un peu après 11 heures du matin on place 10 billes dans l'urne. Une demi-heure plus tard (un peu après 11 heures 30), on en retire une. Une fois passé un quart d'heure, on en ajoute 10 nouvelles. On attend 1/8 d'heure et on retire une bille. On attend 1/16 d'heure et on en ajoute 10 nouvelles. On attend 1/32 d'heure et on en retire une. Etc. On imagine ici qu'on est capable d'agir de plus en plus vite sans limitation. À midi toutes les étapes se sont déroulées car $1/2 + 1/4 + \cdots + 1/2^n + \cdots = 1$. On se demande combien y a-t-il de billes dans l'urne à midi ?

La réponse qui vient immédiatement à l'esprit est qu'il y en a une infinité car on en ajoute beaucoup plus qu'on en retire. Soyons précis. On numérote les instants délimitant nos actions $t_0 = 11\,\text{h}$, $t_1 = 12\,\text{h} - 1/2\,\text{h}$, $t_2 = 12\,\text{h} - 1/4\,\text{h}$, $t_3 = 12\,\text{h} - 1/8\,\text{h}$, …, $t_n = 12\,\text{h} - 1/2^n\,\text{h}$, … Le nombre de billes dans l'urne à l'instant t_n est donné par le tableau :

Instant numéro n	0	1	2	3	4	5	6	…	$2k$	$2k+1$	…
Nombre de billes	0	10	9	19	18	28	27	…	$9k$	$9k+10$	…

1. Professeur émérite, université de Lille, laboratoire CRISTAL UMR CNRS 9189.

DOI:10.48556/SIF.1024.24.207

Il ne fait aucun doute que le nombre de billes présentes dans l'urne tend vers l'infini quand n tend vers l'infini, c'est-à-dire quand t tend vers midi. Il y a donc une infinité de billes dans l'urne à midi.

Pas si vite, ce n'est pas certain ! Imaginons en effet qu'on numérote les billes de 1 à 10 pour les 10 premières déposées dans l'urne, de 11 à 20 pour les 10 suivantes, etc. Imaginons aussi qu'à chaque fois qu'on retire une bille de l'urne on prenne celle de plus petit numéro parmi celles encore présentes. La bille numéro 1 est donc retirée avant l'instant 2, la bille numéro 2 est retirée avant l'instant 4, la bille numéro 3 est retirée avant l'instant 6, etc. Chaque bille déposée dans l'urne finit par être retirée… et donc à midi, il n'y a aucune bille dans l'urne. C'est en contradiction avec l'autre raisonnement. Nous sommes face à un paradoxe !

Pour départager les deux options, on considère qu'il est plus raisonnable d'imaginer qu'à chaque retrait, on choisit la bille au hasard équitablement parmi les billes présentes. Que se passe-t-il ?

Solution

Merci et bravo à Éric Wegrzynowski qui a parfaitement résolu l'énigme. Commençons par considérer une quelconque des dix premières billes.

- La probabilité qu'elle n'ait pas été retirée à l'instant t_2 est $p_1 = 9/10$ car il y avait 10 billes à l'instant t_1, et qu'entre t_1 et t_2, on en a retiré une au hasard.
- La probabilité qu'elle n'ait pas été retirée avant t_4 est $p_2 = (9/10)(18/19)$ car il faut à la fois qu'elle soit restée entre t_1 et t_2, et entre t_3 et t_4. Etc.
- La probabilité qu'elle n'ait pas été retirée avant l'instant t_{2n} est :

$$p_n = (9/10)(18/19)\dots(9n/(9n+1)) \tag{1}$$

La probabilité p_n tend vers 0 quand n tend vers l'infini, car

$$1/p_n = (10/9)(19/18)\dots(9n+1)/9n = (1+1/9)(1+1/18)\dots(1+1/9n) \tag{2}$$

En développant : $1/p_n > 1/9 + 1/18 \cdots 1/9n = (1 + 1/2 + \cdots + 1/n)/9$. Au numérateur la série harmonique tend vers l'infini ; donc p_n tend vers 0. La probabilité est donc nulle à midi, ce qui signifie que la bille considérée est retirée avant midi.

Le raisonnement s'étend à chacune des billes mises à un moment ou un autre dans l'urne, et donc dans le cas d'un retrait au hasard, à midi il ne reste aucune bille dans l'urne comme lorsqu'on retire la bille de plus petit numéro.

Sur un plan mathématique, il n'y a pas de paradoxe car la première conclusion se base sur l'idée qu'à midi le nombre de billes dans l'urne est la limite du nombre de billes dans l'urne à l'instant t_n, or c'est une conclusion mathématiquement infondée. Expliquons pourquoi.

La limite d'une suite E_n d'ensembles existe si, par définition, la limite supérieure des E_n (les éléments qui sont dans une infinité de E_n) est égale à la limite inférieure des E_n (les éléments qui sont dans tous les E_n à partir d'un certain rang). Or quand la limite des E_n existe, il se peut très bien que le nombre d'éléments de E_n soit n et que la limite des E_n soit vide. C'est le cas en prenant $E_n = \{n+1, n+2, \ldots, 2n\}$. Sur le plan purement mathématique le premier raisonnement par décompte est donc faux car la limite des cardinaux des E_n n'est pas égale en général au cardinal de la limite des E_n quand cette limite existe.

Sur un plan mathématique tous les cas sont possibles selon la procédure précise d'ajout et de retrait même en respectant {ajout de 10 billes} $-$ {retrait de 1 bille}, alternativement. En effet la limite d'une suite d'ensembles E_n, où E_{n+1} est obtenu en ajoutant dix entiers à E_n quand n est pair et en en retirant 1 quand n est impair, peut être n'importe quel sous-ensemble fini ou infini de l'ensemble des entiers (démontrez-le !).

Du côté de la physique la méthode pour se défaire du paradoxe est plus facile : la situation envisagée n'est pas réalisable car il faudrait opérer des mouvements de plus en plus rapides et il faudrait donc à un moment aller plus vite que la lumière… ce qui est impossible d'après la théorie de la relativité. Une autre impossibilité physique provient de ce qu'il n'est pas envisageable d'utiliser une infinité de billes, ou d'avoir à sa disposition une urne pouvant contenir une infinité de billes, ou encore de disposer d'assez d'énergie pour opérer les déplacements des billes en moins d'une heure.

Nouveau problème. Une bataille sans bataille

On dispose d'un jeu de cartes qui en contient n n un entier supérieur ou égal à 2. Chaque carte porte un numéro de 1 à n qui détermine sa force. On mélange les cartes, on coupe le paquet en deux sous-paquets qui peuvent être inégaux. On tire à pile ou face l'attribution des paquets. Les joueurs tiennent leur paquet les faces vers le sol (donc sans voir la force des cartes). Chacun pose la carte du dessus de son paquet sur la table en la retournant. La carte la plus forte gagne. Le joueur gagnant range les deux cartes jouées sous son paquet en plaçant la sienne en premier puis celle qu'il vient de prendre. Il place la sienne en premier car elle est la plus forte des deux et qu'il souhaite qu'elle revienne le plus rapidement possible. Il n'y a jamais de bataille (deux cartes de même force) car ici toutes les cartes ont des forces différentes. Si l'un des joueurs n'a plus de cartes, il a perdu. Tout est donc parfaitement déterminé dès que la distribution des cartes est fixée.

La question est : se peut-il qu'une partie soit infinie, c'est-à-dire qu'elle boucle ? Plus précisément, pour quelles valeurs de n et quelles distributions a-t-on une boucle ?

On peut aborder le problème en faisant des programmes (qui vont rechercher des cycles) ou par le raisonnement. Précisons que le problème n'est pas facile et que d'ailleurs, il est aujourd'hui incomplètement résolu.

Envoyez vos réponses à jean-paul.delahaye@univ-lille.fr. Les noms des premiers lecteurs à me donner la bonne réponse (et à la justifier) seront mentionnés dans le prochain numéro de 1024.

Défi c0d1ngUP Chiffrement Cybermen-Daleks

c0d1ngUP team[1]

c0d1ngUP[2] est un **challenge de programmation**, qui s'adresse autant aux débutants qu'aux programmeurs plus expérimentés. Il se présente sous la forme d'épreuves et d'énigmes dont la résolution nécessite **programmation, réflexion et inventivité.** Chaque participant **utilise les outils et langages de programmation de son choix :** seule importe la résolution effective de l'énigme.

La première édition a eu lieu en 2014. Exception faite de 2020, il y a eu une édition par an, généralement en mars. Le challenge se déroule en présentiel, à Poitiers, sur une journée. Cependant, depuis 2021, il est également possible de participer au concours en mode distanciel (quelques défis en moins, tableau des scores adapté…). Le niveau des épreuves est varié, et chaque participant peut, indépendamment de son niveau, trouver des énigmes à résoudre tout en s'amusant.

Les participants peuvent concourir **individuellement ou en équipe de deux** et sont segmentés en différentes catégories qui n'entrent pas en concurrence. Durant la journée, chaque concurrent peut suivre en direct le nombre de points et le classement de tous les participants du concours. Après 7 heures d'intense activité, les résultats sont donnés et les meilleurs de chaque catégorie récompensés.

1. Emmanuel Laizé (université de Poitiers, UP), Freddy Légé (UP), Laurent Signac (UP), Benoit Tremblais (UP), Loïc Restoux, Caroline Tartary.
2. https://codingup.fr.

DOI:10.48556/SIF.1024.24.211

c0d1ngUP poursuit plusieurs objectifs : **promouvoir et démystifier la programmation** auprès du plus grand nombre, **informer les futurs étudiants** sur les filières universitaires qui ouvrent sur les métiers du développement et enfin, permettre aux personnes intéressées par la programmation de **se rencontrer et de mettre à l'épreuve leur savoir-faire.**

Les problèmes les plus faciles peuvent être résolus par des collégiens avancés en programmation (à l'aide d'outils de programmation par blocs comme Scratch). Les plus difficiles donnent du fil à retordre à la plupart des challengers (le temps est limité, et il y a généralement entre 20 et 25 problèmes).

Dans cette nouvelle rubrique, nous proposerons aux lecteurs de 1024 certains de ces problèmes, plutôt parmi les difficiles (donc *a priori* hors de portée des collégiens). Les énoncés seront présentés avec seulement quelques modifications cosmétiques mineures. Des liens seront fournis pour accéder aux fichiers parfois nécessaires à la **résolution effective** du problème (on rappelle que les challengers ne doivent pas produire leur code, mais seulement résoudre effectivement le problème). Dans le bulletin 1024 suivant, nous donnerons quelques *éléments de résolution* agrémentés si possible des idées ou programmes (fonctionnels) que vous aurez pu nous faire parvenir.

Nous présentons ici un problème de 2023, l'année du thème *Docteur Who*[3]. C'est inhabituel, l'énoncé de ce challenge contient aussi du code C.

Chiffrement Cybermen-Daleks

Le Docteur a besoin d'un cryptanalyste, et c'est à vous qu'il a fait appel. Il semble que les Cybermen et les Daleks utilisent un nouveau système de chiffrement pour communiquer. Vous avez pu mettre la main sur la méthode de chiffrement dans un dépôt GitInvaders et vous disposez maintenant de fonctions de chiffrement écrites en `C` :

```c
#include <string.h>
#include <assert.h>
#define BLKSIZE1 5
#define BLKSIZE2 5
#define MODULE 33554432L

char alp[] = "ABCDEFGHIJKLMNOPQRSTUVWXYZ. ,:!?";

unsigned long int pgcd(unsigned long int a,
                       unsigned long int b) {
    if (b == 0) return a;
    return pgcd(b, a % b);
}
```

3. Chaque année a son thème, souvent tiré de la pop culture, et avec lequel nous prenons généralement quelques libertés… que les fans absolus ne s'en offusquent pas.

```
int letter_num(char c) {
    switch(c) {
      case '.': return 26;
      case ' ': return 27;
      case ',': return 28;
      case ':': return 29;
      case '!': return 30;
      case '?': return 31;
      default: return c - 'A';
    }
}
char num_letter(int n) {
    return alp[n];
}

void crypt1_blk(char blk[], char key[], char res[]) {
    for (int i = 0; i < BLKSIZE1; i++) {
        res[i] = num_letter(letter_num(blk[i]) ^
                            letter_num(key[i]));
    }
}
void crypt2_blk(char blk[], unsigned long int key,
                char res[]) {
    unsigned long int n = 0;
    for (int i = 0; i < BLKSIZE2; i++) {
        n = n * 32 + letter_num(blk[i]);
    }
    n = (n * key) % MODULE;
    for (int i = BLKSIZE2 - 1; i >= 0; i--) {
        res[i] = num_letter(n % 32);
        n = n / 32;
    }
}
void crypt(char message[], char key1[], int key2,
           char res[]) {
    assert (strlen(message)
            % (BLKSIZE1 * BLKSIZE2
               / pgcd(BLKSIZE1, BLKSIZE2)) == 0);
    assert (pgcd(key2, MODULE) == 1);
    assert (strlen(key1) == BLKSIZE1);
    assert (key2 >= 0 && key2 < MODULE);

   for (int i = 0; i < strlen(message); i += BLKSIZE1) {
       crypt1_blk(&message[i], key1, &res[i]);
   }
   for (int i = 0; i < strlen(message); i += BLKSIZE2) {
       crypt2_blk(&res[i], key2, &res[i]);
   }
}
```

Les Daleks et les Cybermen ont certainement une paire de clés communes (`key1` et `key2` du code C) et chiffrent leurs messages avec ces clés.

Le Tardis a aussi pu intercepter deux messages :

– Cybermen aux Daleks (le message fait une seule ligne de 70 caractères, les blancs y sont représentés par ce symbole « ␣ ») :

```
:RMOP,DOCC,QTUY,QMHWVA␣L␣DCUUGYMHN␣QUQXWK!
KJGL?FRD␣EL,OJQMOXUIYSDOBMQF
```

– Réponse des Daleks (10 caractères) : `MA.AYTBUAY`

Il y a fort à parier que la dernière réponse des Daleks est : `EXTERMINER`

Serez-vous capable de décrypter le message et d'informer le Docteur du plan machiavélique qui se trame ? Validez le défi avec le message des Cybermen.

Vous pouvez nous faire parvenir votre méthode, vos astuces de résolution, vos remarques, accompagnées de la solution au défi (ici, la réponse au défi est simplement le message décrypté des Cybermen), à l'adresse suivante : mailto:codingup.team+1024@gmail.com. Dans le prochain bulletin, nous proposerons une synthèse de notre solution et de celles qui nous seront parvenues à temps.

Données accessibles en ligne

– Code C de la méthode de chiffrement[4]
– Message des Cybermen[5]
– Réponse des Daleks[6]

4. https://codingup.socinfo.fr/1024/24/pb/chiffrement.c.
5. https://codingup.socinfo.fr/1024/24/pb/message1.txt.
6. https://codingup.socinfo.fr/1024/24/pb/message2.txt.

La SIF remercie ses soutiens pour leur participation à la diffusion de 1024.

Le Cigref

Le Cigref est une association professionnelle à but non lucratif et sans activité commerciale, dont les 157 membres sont les grandes entreprises françaises et les grandes administrations publiques françaises utilisatrices de produits et services numériques. Les membres du Cigref sont représentatifs de l'économie française et des différents secteurs d'activité qui la structurent (banque, assurance, énergie, distribution, transports, industrie, services, administration…). Les entreprises et administrations publiques membres du Cigref sont représentées par leur dirigeant numérique et technologique (DNum, CIO, CTO, CDO, CITO, CDIO, etc.) La gouvernance du Cigref est assurée par un conseil d'administration de 15 membres élus par l'assemblée générale des représentants des membres du Cigref.
Les 157 membres du Cigref représentent environ :

– 2 000 milliards d'euros de chiffre d'affaires cumulé ;
– 10 millions de salariés servis par les prestations de leurs directions du numérique ;
– 400 000 praticiens du numérique ;
– 70 milliards d'euros de cash out annuel dans l'IT.

Les produits et services numériques développés et déployés par les membres du Cigref comptent, chaque jour, et sur toute la planète, des centaines de millions d'utilisateurs, comme clients, bénéficiaires ou partenaires, à titre personnel ou professionnel.

La raison d'être du Cigref est d'agir au service de ses adhérents et de l'intérêt général afin de bâtir le numérique durable, responsable et de confiance que nous voulons collectivement pour la société et son économie. Sa mission pour y parvenir est de développer la capacité des grandes entreprises et administrations publiques à comprendre, intégrer et maîtriser le numérique. Pour réussir sa mission, l'activité du Cigref s'articule autour de trois axes principaux, qui font sa singularité : l'appartenance, l'intelligence collective et l'influence.

CNRS Sciences informatiques

CNRS Sciences informatiques contribue à développer, coordonner et faire connaître les recherches dans son périmètre scientifique, qui va de l'algorithmique et du calcul à la robotique et à l'interaction humain-machine, en passant par les sciences du logiciel, l'intelligence artificielle, la science des données et le traitement du signal, des images, des langues et de la parole. Ces recherches se développent en interaction avec les autres sciences (sciences du vivant et du climat, sciences humaines et sociales, ingénierie, technologies quantiques, etc.) et avec des grandes questions de société (sécurité, sobriété, santé, éducation, etc.). Elles donnent souvent naissance à des innovations, au bénéfice du monde socio-économique et culturel.

CNRS Sciences informatiques soutient les travaux de chercheurs et chercheuses engagés dans ces domaines. Avec ses partenaires académiques, il pilote une cinquantaine de laboratoires en France et à l'étranger ainsi que plusieurs programmes de recherche financés par France 2030. Il contribue à la structuration du paysage scientifique national en animant près d'une vingtaine de réseaux thématiques (GDR) et en opérant des infrastructures de recherche, notamment le supercalculateur Jean Zay.

Inria

Inria est l'institut national de recherche en sciences et technologies du numérique et a la responsabilité depuis janvier 2024 de l'Agence de programmes dans le numérique (algorithmes, logiciels et usages) pour renforcer les dynamiques collectives de la recherche française. La recherche de rang mondial, l'innovation technologique et le risque entrepreneurial constituent son ADN. Au sein de 220 équipes projets, pour la plupart communes avec les grandes universités de recherche, plus de 3 800 scientifiques y explorent des voies nouvelles, souvent dans l'interdisciplinarité et en collaboration avec des partenaires industriels pour répondre à des défis ambitieux. Institut technologique, Inria soutient la diversité des voies de l'innovation : de l'édition *open source* de logiciels à la création de startups technologiques (Deeptech). Enfin, Inria promeut depuis plus de 20 ans des actions de médiation scientifique, pour et par le numérique, notamment vers les scolaires et plus particulièrement les jeunes filles, et œuvre fortement pour la science ouverte.

Moteur de la recherche et de l'innovation numérique en France et en Europe depuis plus de 50 ans, l'objectif d'Inria est de relever le défi de la compétition en développant notamment une politique de site explicite avec les universités françaises intensives en recherche pour renforcer l'attractivité internationale de ces sites universitaires, leur leadership … et leur impact scientifique, technologique et industriel dans les domaines des algorithmes et de l'informatique quantique, du calcul haute performance, de l'éducation et du numérique, de l'intelligence artificielle, de l'Internet des objets, du logiciel, de la modélisation et de la simulation, du numérique et de l'environnement, de la robotique, de la santé numérique, et des sciences des données.

La fondation ANTHONY MAINGUENÉ

Anthony Mainguené, ingénieur informaticien et spécialiste des réseaux et de la cyber-sécurité, était le responsable de la sécurité technique des systèmes d'information au sein de la direction du groupe Bouygues construction à Challenger (Guyancourt). Il est décédé brutalement d'une hémorragie cérébrale à l'âge de 40 ans. Ses compétences et sa conception managériale favorisant le bien-être et l'efficacité au travail, son intégrité et son sens humain, étaient reconnus de tous.

La fondation Anthony Mainguené, à vocation universelle, a été créée afin de transmettre son sens de l'éthique, ses idées et ses espérances. Elle est placée sous l'égide de la Fondation de France et est soutenue par le conseil régional de la Nouvelle Aquitaine.

Le monde change et la place de l'homme avec lui. La fondation souhaite faire prendre conscience que la réflexion éthique est devenue incontournable et ne peut se réaliser pleinement que dans et par l'action.

L'éthique permet de s'interroger et d'agir en toute responsabilité, avec engagement, solidarité, équité et respect de l'Autre.

Pour ce faire, la fondation intervient selon 5 axes : le droit, le numérique, l'ingénierie, les sciences humaines et sociales, le développement durable.

La fondation Anthony Mainguené promeut les prises de conscience éthiques au moyen de colloques, séminaires, enseignements, remises de prix, et s'attache à réussir le challenge de former des Femmes et des Hommes responsables qui sachent innover et œuvrer en toute liberté de conscience pour relever les grands défis d'une société en mutation.

« *L'humanité est une aventure solidaire et responsable qui passe par l'autre.* »

Anthony Mainguené

Le syndicat Numeum

Numeum est le syndicat et l'organisation professionnelle de l'écosystème numérique en France. Il représente les entreprises de services du numérique (ESN), les éditeurs de logiciels, les plateformes et les sociétés d'ingénierie et de conseil en technologies (ICT). Numeum rassemble plus de 2 500 entreprises adhérentes qui réalisent 85 % du chiffre d'affaires total du secteur. Présidé par Véronique Torner, Numeum se fixe trois priorités : les territoires, pour accompagner les adhérents en région, les compétences, pour répondre aux défis de la mixité et de l'attractivité, et le numérique responsable pour accompagner et soutenir le développement d'un écosystème numérique dans une trajectoire d'impact positif sur le plan économique, social, sociétal et environnemental à l'échelle européenne, nationale et locale.

Numeum est membre de la fédération Syntec. Le secteur du numérique représente 70 milliards d'euros de chiffre d'affaires et 670 000 employés en France. `http://www.numeum.fr`.

POURQUOI ADHÉRER À LA SIF ?

01.
Rejoindre une communauté ouverte sur le monde

02.
Développer son réseau

03.
Bénéficier d'un accès privilégié aux manifestations proposées par la SIF et ses partenaires

04.
Recevoir le bulletin 1024 dès sa parution

Adhérer c'est aussi

- Contribuer à la légitimité de la SIF en augmentant sa représentativité
- S'engager pour une informatique plus juste, plus éthique, plus responsable

REJOIGNEZ-NOUS !
adhesion.socinfo.fr

FAITES UN DON

Particuliers **Entreprises**

Soutenez nos actions en faisant un don **défiscalisé** à la SIF

don.socinfo.fr

Avec le soutien de

Cigref
RÉUSSIR
LE NUMÉRIQUE

cnrs

Inria

FONDATION ANTHONY MANGUERRE
Compétence Conscience Cœur
Promouvoir des prises de consciences éthiques
Sous l'égide de la Fondation de France

numeum

Sif
Société
informatique de France

Institut H. Poincaré · 11 rue Pierre & Marie Curie · 75231 PARIS Cedex 05