



GOVERNEMENT

*Liberté
Égalité
Fraternité*

**Pôle d'expertise de la
régulation numérique – PEReN**

Rapport d'activité 2022

MINING

**ARTIFICIAL
INTELLIGENCE**

**PROBLEM
SOLVING**

AUTOMATION

**PATTERN
RECOGNITION**

**MACHINE
LEARNING**

ALGORITHM

Sommaire

Le Pôle d'expertise de la régulation numérique	5
Édito	6
Les temps forts 2022	7
Les chiffres clés 2022	8
Partie 1	
Accompagner la préparation et la négociation des textes	10
Partie 2	
Doter les administrations d'outils de régulation et d'évaluation	15
Partie 3	
Déployer des ressources mutualisées	20
Partie 4	
Travailler à l'état de l'art et diffuser les savoirs	24
Partie 5	
Nouveaux pouvoirs et montée en puissance	29

Note à l'intention du lecteur

Du fait d'un niveau de confidentialité qui peut parfois être très élevé, certains projets sur lesquels travaille le PEReN n'apparaissent pas sur ce rapport ou ne peuvent pas être décrits précisément.

PEReN

Pôle d'Expertise de la Régulation Numérique

Mission

Le PEReN offre, à l'ensemble des administrations d'État intervenant dans la régulation des plateformes numériques, une expertise et une assistance technique dans le domaine des sciences des données et du numérique. Il fournit ainsi des outils, des études et du conseil entièrement sur ses ressources propres. Il s'investit également dans des projets de recherche à caractère exploratoire ou scientifique.

Objectifs

- Renforcer l'efficacité de l'action publique en permettant l'accès à une force de frappe mutualisée ;
- Mieux appréhender le fonctionnement des plateformes numériques ;
- Accompagner la définition de la réglementation ;
- Favoriser le partage des savoirs.

Statut

Service à compétence nationale, le PEReN est placé sous l'autorité conjointe des ministres chargés de l'économie, de la communication et du numérique, avec une vocation interministérielle et au service des régulateurs indépendants.

Équipe

Le PEReN mobilise une équipe d'experts « *data scientists* » et développeurs de haut niveau, docteurs ou ingénieurs, aux compétences techniques rares en matière d'analyse de données, de code source, de traitements et d'audit algorithmiques utilisés par les plateformes numériques.



Édito



NICOLAS DEFFIEUX
Directeur du PEReN

Que retenir de l'année 2022 pour le PEReN ?

2022 est une année charnière de consolidation. Deuxième année de plein exercice, elle s'est déroulée au rythme de plus d'une

soixantaine de projets à mener, soit 3 fois plus qu'en 2021. Cette activité en forte croissance témoigne de la reconnaissance du PEReN au sein de l'écosystème du numérique. Son expertise et son utilité font consensus auprès des acteurs qui identifient le Pôle de plus en plus spontanément. Le Conseil d'État, dans son étude annuelle consacrée aux réseaux sociaux, atteste d'ailleurs de cette réussite. Cette reconnaissance a même acquis une dimension européenne ! La Commission a retenu un modèle similaire au PEReN pour configurer son Centre européen pour la transparence algorithmique (ECAT), chargé de lui fournir un appui technique dans le cadre des récents règlements promulgués pour la régulation des services numériques européens.

Quels travaux ont singulièrement marqué cette année ?

2022, c'est également un élargissement du cadre d'intervention du Pôle. Le législateur a étendu fin 2021 aux autorités indépendantes (Arcom, Autorité de la Concurrence, CNIL, Défenseur des droits...) la liste des administrations pouvant faire appel au PEReN et l'a doté de nouveaux pouvoirs pour la conduite d'expérimentations et de travaux de recherche publique. Dès 2022, toutes les autorités indépendantes à qui cette possibilité était offerte ont confirmé leur demande d'association. En parallèle des projets conduits pour nos partenaires, de premières expérimentations destinées au prototypage d'outils utiles à leurs missions de régulation ont

été conduites et ont fait l'objet d'un rapport annuel d'évaluation remis au Parlement et à la CNIL au printemps 2023. Nous avons investi le champ de la recherche académique au travers de projets dédiés et présentés à la communauté, de l'accueil d'un doctorant en collaboration avec Inria, le CNRS et l'Université de Rennes et de notre inscription comme structure de recherche au registre national (RNSR).

Quelle est la feuille de route du PEReN pour 2023 ?

Dès l'origine, les missions et la feuille de route confiées au PEReN faisaient preuve d'une grande ambition pour un modèle entièrement à inventer. Cette ambition doit être maintenue, les champs du numérique restant à investir sont nombreux. Au-delà du fonctionnement des plateformes numériques, l'expertise de fond allée à une maîtrise de première main des techniques et des enjeux de l'intelligence artificielle acquise au PEReN est à thésauriser et réexploiter sur les champs ouverts par les modèles génératifs et des giga-modèles de langage (LLM).

Notre phase d'installation est achevée mais les besoins qui nous sont adressés continuent de croître : le chantier reste ouvert pour accompagner les régulateurs et leurs équipes dans le recours et la prise en main des infrastructures et outils régulièrement mis à disposition par le Pôle.

Qu'il apporte une expertise destinée aux décideurs ou équipes responsables de l'élaboration des réglementations, au niveau européen ou national, ou qu'il apporte un appui direct aux régulateurs dans les aspects les plus techniques de la mise en œuvre de cette régulation, le PEReN a ceci de singulier d'être tout autant capable de faire que d'analyser. L'équipe qui le constitue est ainsi en mesure de concevoir des outils, développer des modèles d'intelligence artificielle et participer à des travaux de recherche au service des pouvoirs publics et de l'intérêt général.

Les temps forts 2022

Février

18/02 > Parution de l'Éclairage sur...n° 2 – « Applications mobiles : quels enjeux de sécurité pour leur distribution hors des magasins des OS ? »

Avril

07/04 > Parution de l'Éclairage sur...n° 3 – « Privacy Sandbox : une collection d'outils pour une publicité en ligne exempte de cookies tiers »

21/04 > Parution du décret¹ d'application de la loi du 25/10/2021 étendant la liste des administrations pouvant faire appel au PEReN et le dotant de nouveaux pouvoirs d'expérimentations et de recherche publique.

Mai

20/05 > Parution de l'Éclairage sur...n° 4 – « Détection des mineurs en ligne : peut-on concilier efficacité, commodité et anonymat ? »

24/05 > Cycle de séminaires académiques du PEReN – « L'enfermement et les élections, par le prisme du recommandeur de YouTube » (Erwan Le Merrer, Gilles Trédan)

Juin

10/06 > Inscription du PEReN au Registre national des structures de recherche

13/06 > Intervention à la 12^e édition de l'Atelier sur la Protection de la Vie Privée organisé par l'équipe R3S du laboratoire SAMOVAR

23/06 > Réunion multilatérale semestrielle avec les partenaires

30/06 > Lancement des 1^{res} expérimentations dans le cadre des nouvelles compétences du PEReN

Juillet

01/07 > Homologation de l'architecture du système d'information du PEReN

Septembre

16/09 > Matinée « Démonstrations du PEReN » à l'intention des partenaires

19/09 > Cycle de séminaires académiques du PEReN – « Applications mobiles, fer de lance du capitalisme de surveillance » (Esther Onfroy)

Octobre

17/10 > Parution de l'Éclairage sur...n° 5 – « AMP de Google : du dopage au sevrage ? »

Novembre

18/11 > Réunion multilatérale semestrielle avec les partenaires

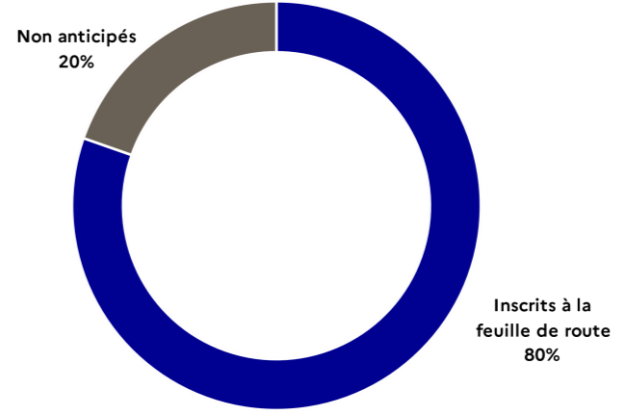
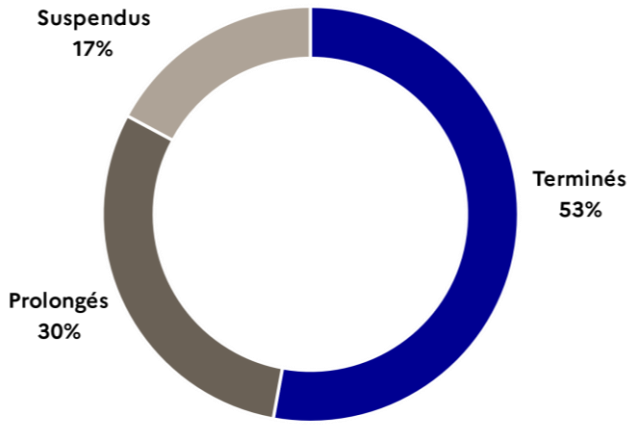
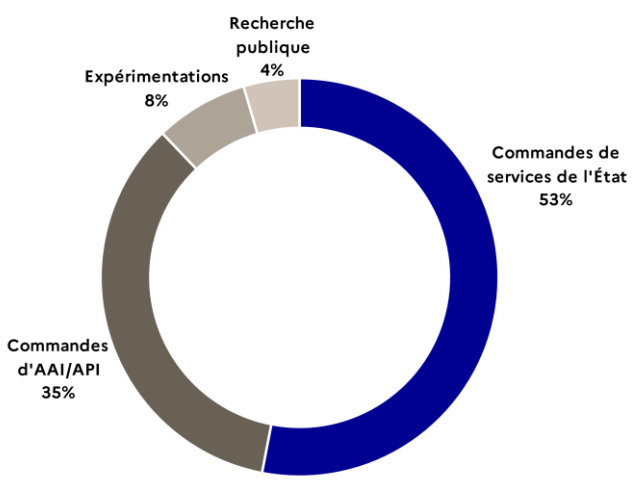
30/11 > Conférences et hackathon « Possibilités technologiques, limites juridiques : les algorithmes face à la régulation », 1^{er} événement annuel du PEReN organisé avec l'appui de la DGE et ses partenaires Inria, le CNNum (Conseil national du numérique) et le département d'économie de Sciences Po Paris

¹ Décret n° 2022-603 du 21 avril 2022 fixant la liste des autorités administratives et publiques indépendantes pouvant recourir à l'appui du pôle d'expertise de la régulation numérique et relatif aux méthodes de collecte de données mises en œuvre par ce service

dans le cadre de ses activités d'expérimentation (<https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000045638527>)

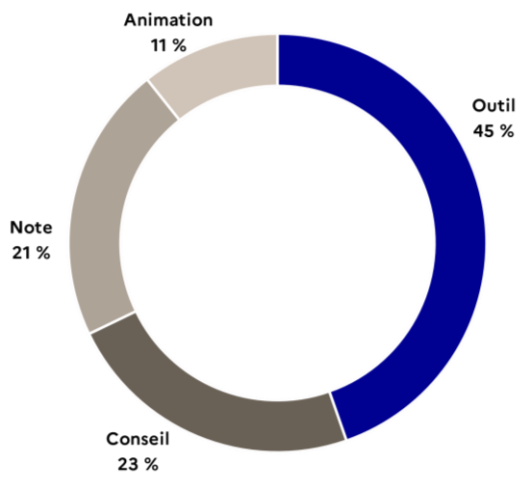
Les chiffres clés 2022

70 projets menés...

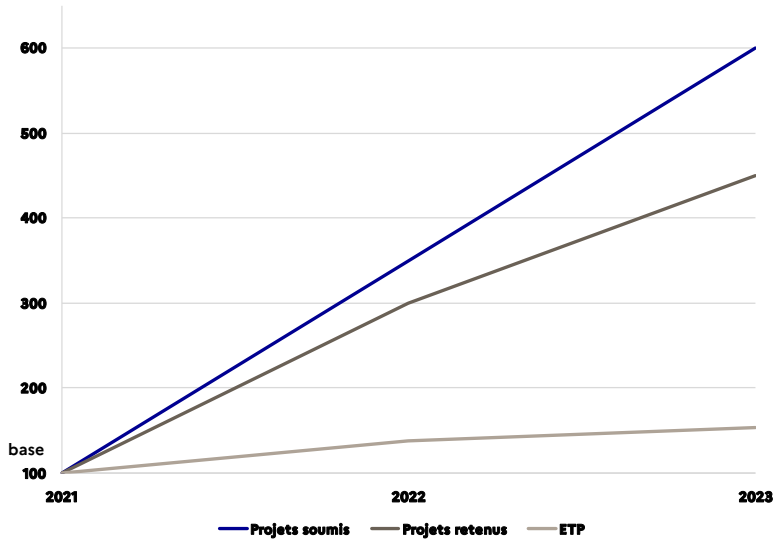


suspendus : en concertation avec les partenaires (dépriorisation, cadrage juridique...)
prolongés : en concertation avec les partenaire ou du fait d'une programmation pluriannuelle

... 56 livrables planifiés



et des effectifs en pleine croissance.



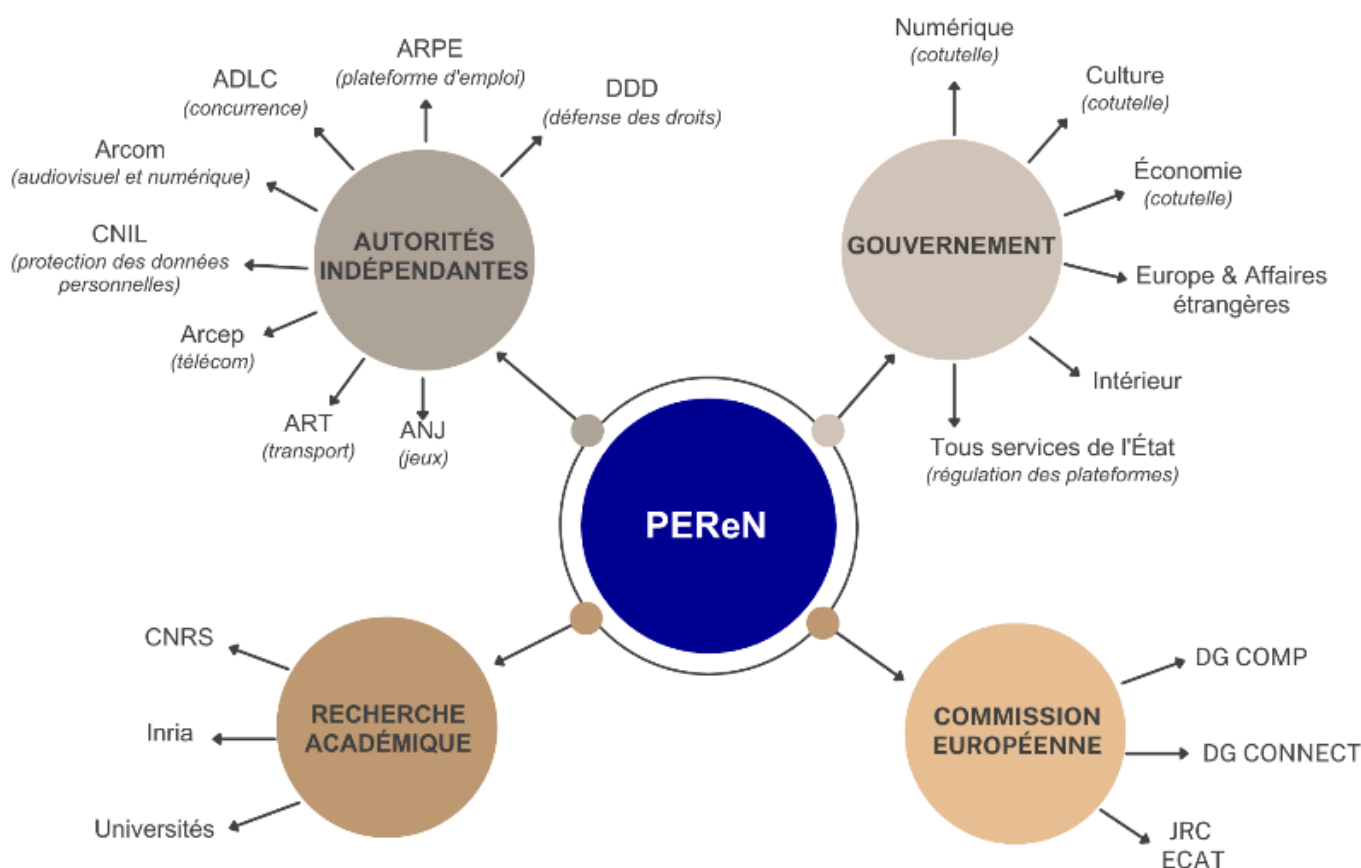
Évolution comparée des volumes de projets soumis, retenus et des ETP, base 100 en 2021

Les moyens du PEReN

Rattaché au Directeur Général des Entreprises pour sa gestion administrative et financière, bénéficiaire à sa création de dotations en provenance du Fonds de transformation ministériel (FTM) et du programme France Relance (guichet ITN5), le PEReN fonctionne aujourd'hui sans ligne budgétaire consacrée et avec des frais de personnel et de fonctionnement pris en charge par la Direction générale des entreprises. Installé depuis bientôt trois ans, le PEReN sort de sa phase d'expérimentation et d'incubation et est devenu une ressource mutualisée dont le besoin fait consensus.

Sa montée en puissance rapide et sa capacité à attirer des profils de haut niveau ont confirmé sa pertinence et l'intérêt de capitaliser au sein de l'État ces compétences. Ce positionnement, qui fait figure de précurseur, devrait aujourd'hui trouver à s'inscrire durablement dans le paysage administratif en le dotant de ressources propres garantissant son activité.

Les partenaires du PEReN



Partie 1

Accompagner la préparation et la négociation des textes

La place désormais prépondérante des plateformes numériques dans notre quotidien appelle une régulation ajustée de cet écosystème où l'innovation perpétuelle est une marque de fabrique. Dans un contexte de forte structuration de cette régulation tant au plan européen que national, le PEReN met à disposition des pouvoirs publics et des régulateurs une expertise technique objective et à l'état de l'art. Son accompagnement s'effectue à tous les stades de la construction de la réglementation, de sa préparation à sa mise en œuvre.

D'un accord européen sous impulsion française à une déclinaison nationale de la régulation de l'espace numérique

L'année 2022 a été marquée par un élan soutenu de la régulation numérique, impulsé notamment au niveau européen par les négociations autour de projets de règlement européen, le *Digital Markets Act* (DMA)² et le *Digital Services Act* (DSA)³. Entrés en vigueur fin 2022, après une adoption sous présidence française du Conseil de l'Union européenne, ces textes, à l'impact majeur sur le secteur numérique, ont pour objectif de doter l'Union européenne d'un nouveau cadre de responsabilité des grandes plateformes numériques tant dans leur dimension sociétale (lutte contre la dissémination des contenus illicites ou préjudiciables) que dans leur dimension économique et concurrentielle (garantir que les marchés restent ouverts à la concurrence et le lieu de pratiques équilibrés et loyales).

Poursuivant son accompagnement technique des services du ministère de l'Économie, des finances et de la souveraineté industrielle et numérique dans les dernières phases de négociations de ces textes, le PEReN a assisté ces mêmes services en charge de l'adaptation de cette réglementation en droit français. Il a, en particulier, fourni un appui aux équipes pilotant la préparation du projet de loi « Sécuriser et réguler l'espace numérique », sur des

² Règlement (UE) 2022/1925 du Parlement européen et du Conseil du 14 septembre 2022 relatif aux marchés contestables et équitables dans le secteur numérique et modifiant les directives (UE) 2019/1937 et (UE) 2020/1828 (règlement sur les marchés numériques) (<https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:32022R1925>)

³ Règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (règlement sur les services numériques) (<https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:32022R2065>)

problématiques relatives à la protection des mineurs et en lien avec ses autres projets en la matière (contrôle parental, vérification de l'âge, ...). Ces travaux ont notamment permis de préciser le champ des possibles sur le plan technique pour les différents acteurs (systèmes d'exploitation, applications mobiles, magasins d'applications, navigateurs, moteurs de recherche) en matière de blocage de contenus et de contrôle du temps d'écran.

Un appui a également été fourni sur la thématique de la publicité programmatique afin d'étudier les mesures possibles (interopérabilité des services, ouverture d'inventaires, transparence, ...) sur ce marché en développement constant.

Contrôle de l'âge en ligne : explorer les possibles

De nombreux types de plateformes sont aujourd'hui dans l'obligation de limiter ou d'interdire leur utilisation par les personnes mineures. Des cas d'usage peuvent être par exemple l'inscription sur les réseaux sociaux, la vente d'alcool en ligne, l'accès aux sites pornographiques ou encore l'interdiction prévue par le règlement DSA de tout profilage des mineurs à des fins publicitaires. Relevant de différents cadres juridiques, ces obligations, ne sont pas à l'évidence aujourd'hui appliquées uniformément par les acteurs. Dans ce contexte, il a semblé important de comprendre les différents écosystèmes et leurs contraintes techniques spécifiques, afin de pouvoir accompagner une mise en conformité sans nuire aux intérêts de l'utilisateur, notamment en termes de vie privée.

Cette démarche a trouvé un intérêt auprès d'une multitude d'acteurs qui ont ainsi pu contribuer, chacun dans leur champ d'expertise et de compétences, à sa mise en œuvre.

S'intéressant aux problématiques conjointes de protection de la vie privée et de vérification de l'âge en ligne, le PEReN s'est associé à la CNIL et au professeur Olivier Blazy (LIX, École Polytechnique) pour la conduite de travaux en la matière. S'appuyant sur la consultation d'acteurs du secteur des grands réseaux sociaux dans un premier temps et sur différentes analyses, le PEReN a dressé un panorama critique des solutions de vérification d'âge actuellement déployées. Il reprend l'ensemble de ses observations et conclusions dans le n°4 de sa collection en ligne, « Éclairage sur... » publié en mai 2022⁴. Le constat était alors à cette date que pratiquement aucun service en ligne n'utilisait de procédé permettant de vérifier l'âge de ses utilisateurs dans des conditions de fiabilité et de garanties de respect de la vie privée satisfaisantes. Malgré leur multiplicité, peu de méthodes apparaissaient à la fois faciles à mettre en œuvre, peu contraignantes et respectueuses de la vie privée des utilisateurs, performantes et robustes face à des tentatives de fraude. S'attendant à la question de la transmission de l'information de la preuve d'âge, un dispositif de « double-anonymat » permettant de garantir qu'aucun acteur n'ait accès à la fois à des données personnelles très identifiantes et aux sites consultés par les internautes s'est révélé techniquement possible, bien qu'aucune solution de ce type n'ait été implémentée à ce stade. Dans le cadre de leur collaboration, le PEReN, la CNIL et le LIX ont ainsi développé une preuve de concept⁵. Si des pistes d'améliorations et des variations sont d'ores et déjà identifiées, ce prototype ouvre néanmoins de nouvelles voies vers une meilleure protection des individus lorsqu'une reconnaissance de l'âge est nécessaire sur internet.

⁴ Éclairage sur...n°4 - Détection des mineurs en ligne : peut-on concilier efficacité, commodité et anonymat ? (https://www.peren.gouv.fr/rapports/2022-05-20%20-%20-%20Eclairage-sur-detection-mineurs_FR.pdf)

⁵ Démonstrateur du mécanisme de vérification de l'âge respectueux de la vie privée (<https://linc.cnil.fr/demonstrateur-du-mecanisme-de-verification-de-lage-respectueux-de-la-vie-privée>)

Dans le fil de ces travaux, le PEReN a fourni en 2022 un appui technique à l'ARCOM (Autorité de régulation de la communication audiovisuelle et numérique), autorité en charge de la régulation de certains des acteurs concernés par la vérification d'âge et aux services du Ministre en charge du numérique qui ont porté, avec la Direction générale des médias et des industries culturelles, la position et l'expertise française dans le domaine lors des négociations européennes sur le *Digital Services Act* (DSA), en lien également avec les services de la Commission Européenne.

Ces travaux se poursuivent en 2023, et le PEReN, en coordination avec et sous la conduite de l'ARCOM et la CNIL, continue d'étudier les différentes solutions de vérification d'âge proposées par les acteurs afin de pouvoir les accompagner au mieux au bénéfice de l'utilisateur.

Lutte contre les abus sexuels sur enfants : vers un prochain règlement européen

En mai 2022 la Commission européenne a proposé un nouveau projet de règlement établissant des règles en vue de prévenir et de combattre les abus sexuels sur enfants⁶. Ce règlement prévoit notamment la mise en place de nouvelles obligations pour les fournisseurs de services en ligne, en ce qui concerne le retrait, le blocage ou la détection de contenus pédopornographiques, connus ou nouveaux, hébergés ou échangés par l'intermédiaire de leurs services.

Dans ce cadre, le PEReN a apporté son expertise aux différents ministères investis dans l'établissement des positions françaises et les négociations sur ce projet de texte. Les champs explorés concernaient la faisabilité technique des obligations du règlement, l'analyse des performances des algorithmes et modèles d'intelligence artificielle pouvant être mis en œuvre pour satisfaire l'obligation de détection, les modèles d'attaques possibles et l'impact des obligations sur le chiffrement et sur la sécurité des systèmes d'information. Par ailleurs, dans le cadre d'un groupe de travail dédié, le PEReN a également participé aux échanges techniques avec les experts de la Commission européenne, échangé avec différents acteurs de la lutte contre la pédocriminalité, produit des éléments de synthèse du fonctionnement des algorithmes de reconnaissance ou de classification de contenu à destination des ministères et mis en œuvre des démonstrateurs de contournement ou de détournement de ces algorithmes.

En 2023, le PEReN continue d'apporter son expertise technique sur la proposition de règlement. Les démonstrateurs et analyses techniques produits dans ce cadre sont transposables à d'autres types de modèles d'intelligence artificielle de reconnaissance ou de classification de contenus.

⁶ Règlement du parlement européen et du conseil établissant des règles en vue de prévenir et de combattre les abus sexuels sur enfants (<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A209%3AFIN>)

Intelligence artificielle : appui technique aux discussions sur le projet européen de règlement

Les systèmes d'intelligence artificielle s'intègrent chaque jour davantage dans les usages numériques des citoyens français et européens. Ils sont utilisés pour traduire du texte, pour répondre plus rapidement aux emails ou encore pour trouver sur les plateformes en ligne du contenu correspondant à leurs centres d'intérêts. Toutefois, certains d'entre eux peuvent comporter des risques pour les droits fondamentaux des utilisateurs. Face à l'émergence de ces technologies sur le marché européen, la Commission européenne a proposé en 2021 un projet de réglementation sur l'intelligence artificielle : l'*AI Act*⁷. Le règlement vise à interdire ou à soumettre à certaines obligations les systèmes dont les finalités sont à risque, tout en permettant aux acteurs européens d'innover en matière d'IA.

Ainsi, l'*AI Act* ambitionne de réguler des produits hautement techniques, dont certains sont très récents comme les systèmes d'intelligence artificielle générative capables de créer de nouveaux contenus à partir de données d'entraînement (ChatGPT, Stable Diffusion, etc.). Dans le cadre de ses missions de promotion de l'innovation et de l'entrepreneuriat en France, la DGE a participé aux négociations européennes autour de ce projet de règlement. Ces négociations ont permis au Conseil de l'Union européenne d'aboutir à une orientation générale au début du mois de décembre 2022. Le PEReN a participé activement aux discussions afin d'accompagner la DGE sur les aspects techniques du texte. Des analyses ont ainsi été produites concernant les systèmes d'identification biométrique à distance ou bien les technologies d'IA à usage général dont font partie les grands modèles de langage.

Après avoir amendé la proposition, le Parlement Européen a adopté sa propre version du texte en plénière au mois de juin 2023. Ce vote marque le début des phases de « trilogues » : la présidence espagnole de l'Union européenne souhaite faire émerger un compromis avant la fin de l'année. Le PEReN sera ainsi amené à poursuivre son activité de veille et à fournir de nouvelles analyses techniques aux services de l'Etat jusqu'à l'adoption de la version définitive du règlement.

Dans ce contexte et face à l'évolution rapide de ces modèles et des réflexions relatives à leur régulation, le PEReN a choisi de consacrer son n°6 d'« Éclairage sur... » aux modèles d'IA conversationnels. Paru en avril 2023, « ChatGPT ou la percée des modèles d'IA conversationnels »⁸ invite ainsi, dans un format accessible à des non experts, à une plongée technique au cœur de ces dispositifs pour en saisir les principaux enjeux et limites.

⁷ Proposition de règlement du parlement européen et du conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'union (<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>)

⁸ Éclairage sur...n°6 – ChatGPT ou la percée des modèles d'IA conversationnels (https://www.peren.gouv.fr/rapports/2023-04-06_Eclairage%20sur_CHATGPT_FR.pdf)

Accompagner la préparation et la négociation des textes

Nos projets en 2022

Régulation de l'espace numérique

Enjeux de l'intégration des navigateurs internet dans le champ des obligations du DMA

Dans le cadre de la proposition législative européenne *Digital Markets Act* (DMA) analyse technique, pour le compte du ministère de l'Économie, des finances et de la souveraineté industrielle et numérique (MEFSIN), des conditions dans lesquelles les navigateurs internet peuvent être considérés comme des services de plateforme essentiel.

DMA & DSA, préconisations pour l'implémentation technique de dispositions

Appui technique des équipes du MEFSIN dans le cadre des négociations des projets de règlement européens *Digital Markets Act* (DMA) et *Digital Services Act* (DSA) en matière de périmètre d'application d'obligations (interopérabilité, *self-preferencing*, *risk assessment*) ou de métriques d'évaluation de leur respect par les opérateurs.

Déclinaison française de la réglementation européenne de l'espace numérique

Dans le prolongement de l'adoption du DMA et du DSA, appui technique des services du ministère de l'Économie, des finances et de la souveraineté industrielle et numérique pour l'adaptation en droit français de cette réglementation européenne dans les champs notamment de la protection des mineurs et de la publicité en ligne.

Harmonisation européenne des règles concernant l'intelligence artificielle

Accompagnement technique des équipes du MEFSIN dans l'établissement de la position française sur le projet de règlement européen sur l'Intelligence artificielle.

Sécurisation de l'espace numérique

Lutte contre les contenus terroristes en ligne – Appel de Christchurch et GIFCT

Sur demande du ministère de l'Europe et des affaires étrangères, le PEReN a représenté les autorités françaises pour l'élaboration de la feuille de route de l'Appel de Christchurch sur l'évaluation des algorithmes. Il les représente également au sein du groupe de travail technique du *Global Internet Forum to Counter Terrorism* (GIFCT), pour lequel il a assuré le rôle de coordinateur des organisations gouvernementales.

Règlement européen de lutte contre les abus sexuels sur mineurs : appui technique

Dans le cadre de la proposition de règlement européen contre les abus sexuels sur mineurs, le PEReN fournit un appui technique aux services français en charge du suivi et de la négociation du règlement, tel que le Secrétariat général des affaires européennes. En 2022, il s'est intéressé notamment à la faisabilité et à la performance des solutions techniques nécessaires au respect des obligations de détection proposées par ce règlement, notamment concernant l'articulation entre obligation de détection et préservation du chiffrement.

Technologies d'identification des mineurs : état des pratiques

Dans le cadre d'une réglementation en construction, production d'un panorama et d'une analyse d'efficacité des solutions techniques d'identification des utilisateurs mineurs actuellement mises en place par les plateformes pour adapter leur fonctionnement. Cette étude s'est intéressée également aux solutions alternatives possibles et aux initiatives européennes ou d'autres états membres.

Contrôle parental : modalités techniques de l'application de la loi Studer

Accompagnement technique, dès 2022, des services de la DGE dans la rédaction du décret d'application de la loi Studer visant à renforcer le contrôle parental sur les moyens d'accès à internet. Paru à l'été 2023, ce décret précise notamment les modalités de ce contrôle parental (accessibilité et fonctionnalités en particulier).

Partie 2

Doter les administrations d'outils de régulation et d'évaluation

Réguler les plateformes numériques impose aux pouvoirs publics concernés de disposer des compétences techniques et outils propres à leur permettre d'appréhender efficacement les pratiques des opérateurs du secteur et le respect de leurs obligations. Au fait des innovations fortes et constantes dans ce domaine, le PEReN met à leur disposition cette expertise de pointe et cette « force de frappe » nécessaires.

La visibilité sur les *marketplaces* : un enjeu pour les vendeurs français

Quelques places de marché en ligne concentrent une grande partie du chiffre d'affaires du e-commerce en France. Dans ce contexte, les stratégies des plateformes dans la mise en avant des produits et la représentativité des vendeurs français parmi les volumes de ventes ont vu leur importance croître ces dernières années. Comprendre et décrire les conditions dans lesquelles les vendeurs tiers sont mis en avant ou sont représentés sur ces places de marché est essentiel pour appréhender la dynamique de ces canaux de distribution.

Les services de la DGE, et plus particulièrement les équipes chargées de la prospective, des études et de l'évaluation économiques ont missionné le PEReN en 2022 pour étudier les offres mises en avant sur les plateformes de e-commerce sur un échantillon de produits populaires. Le PEReN a ainsi relevé et analysé automatiquement les informations publiquement disponibles pour des produits référencés sur des plateformes de e-commerce en France, en Allemagne et en Italie. Les données ont été enrichies avec les coordonnées postales mises en ligne sur les plateformes par les vendeurs afin d'estimer leur pays d'origine. L'exploitation de ces données a permis aux services de la DGE d'établir un panorama descriptif de la mise en avant de certains vendeurs par rapport à leurs concurrents et de la typologie des vendeurs présents par secteur sur les places de marché étudiées. Ces analyses permettront de formuler des recommandations de politiques publiques pour accompagner les différents secteurs économiques sur la vente en ligne.

Le projet pourra donner lieu à différents prolongements comme des audits algorithmiques ou une poursuite de l'analyse sur un panier représentatif des biens commercialisés en ligne.

Mesurer la valeur du consentement des usagers dans la publicité en ligne

La publicité en ligne est une des principales sources de financement pour les éditeurs de contenus. À ce titre, elle est souvent citée comme moteur du modèle économique d'un web ouvert et gratuit.

Le secteur de la publicité en ligne peut s'appuyer à l'heure actuelle sur différentes méthodes pour suivre et cibler les annonceurs. En effet, la publicité ciblée est vendue plus chère aux annonceurs que la publicité non-ciblée. Néanmoins, le ciblage s'appuie sur le dépôt de cookies qui sont soumis au consentement des visiteurs. Une question qui est peu abordée est la valorisation de ce consentement par les annonceurs et les paramètres qui jouent sur le montant que l'annonceur est prêt à payer pour afficher sa publicité.

Pour explorer cette question, la CNIL s'est tournée vers le PEReN pour réaliser une analyse exploratoire de l'impact du parcours utilisateur sur les enchères qu'il génère sur les sites d'éditeurs. Le PEReN a donc collecté automatiquement des données d'enchères de publicité en ligne réalisées à l'aide de la technique de *Header Bidding*⁹ pour un échantillon d'éditeurs populaires en France. Ces données ont été récupérées selon différents scénarios : acceptation, refus de cookies, parcours de consultation de plusieurs sites web (luxe, voyage, divertissement, enfant), mise dans le panier de produits.

Ces données ont alimenté le laboratoire d'innovation numérique de la CNIL (LINC) dans le cadre de travaux internes sur les modèles de financement du web.

Fake streams : lutter contre la fraude des fausses écoutes en ligne

À l'heure où la *streaming* (lecture instantanée de contenus audio ou vidéo en ligne, sans téléchargement) représente 70% des revenus de l'industrie musicale française, les artistes comme les *majors* de l'industrie s'inquiètent de l'ampleur prise par la manipulation des écoutes en ligne. En effet, que ce soit pour générer des revenus, obtenir la signature d'un contrat, d'une distinction musicale ou encore pour nuire à un concurrent, les *fake streams* – ou fausses écoutes – viennent altérer la mesure clé de la rémunération et de la réputation d'un artiste : le nombre d'écoutes. En l'absence de contrôles, des artistes ayant recours aux *fake streams* sont susceptibles de voir leur rémunération augmenter au détriment des autres, celle-ci étant calculée au prorata du nombre total d'écoutes sur la plateforme. La confiance entre les acteurs de l'industrie peut s'en trouver affectée.

Le Centre national de la musique (CNM) a été missionné pour réaliser la première étude au monde sur ce sujet, documentée et concertée, qui aura duré 18 mois. Le PEReN, en lien avec la DGMIC (Direction générale des médias et des industries culturelles) est venu en soutien à la production de cette étude en réalisant un état de l'art de la recherche sur la manipulation des écoutes sur les plateformes musicales et en proposant une liste d'indicateurs de transparence actionnables avec les plateformes de *streaming* afin de suivre l'évolution des *fake streams* et les moyens de lutte contre ce phénomène.

⁹ *Header bidding* : méthode qui permet à un éditeur en ligne d'organiser une mise en concurrence des différents réseaux publicitaires lors de la vente de son inventaire et donc de maximiser ses revenus (source CNIL, <https://www.cnil.fr/fr/definition/header-bidding>)

Les indicateurs de transparence produits ont fait l'objet d'une préconisation dans le rapport sur les *fake streams* publié par le CNM le 16 Janvier 2023¹⁰. Ils seront retravaillés dans le cadre de la poursuite des travaux du CNM, en concertation avec les acteurs de l'industrie, courant de l'année 2023.

Recommandation de contenus : quels sont les effets d'un algorithme ne reposant pas sur du profilage ?

L'article 38 du règlement européen sur les services numériques (*Digital Services Act*) impose aux très grandes plateformes de proposer à leurs utilisateurs au moins une option sur chaque algorithme de recommandation de contenus ne reposant pas sur le profilage, autrement dit la collecte de données personnelles à des fins de personnalisation de l'expérience utilisateur.

Pour anticiper les changements liés à l'application de cette nouvelle réglementation, notamment sur la pertinence des contenus proposés ou l'exposition aux contenus préjudiciables, la Direction Générale des Médias et des Industries Culturelles (DGMIC) du ministère de la Culture a demandé au PEReN d'évaluer les conséquences de l'utilisation d'algorithmes de recommandation non-fondés sur le profilage des utilisateurs. Focus particulier pour ce projet : les contenus produits par les éditeurs de presse et les changements de visibilité induits par ces nouveaux algorithmes.

Pour répondre à cette demande le PEReN a, dès 2022, analysé l'état de l'art de la recommandation de contenus pour produire une synthèse du fonctionnement des algorithmes de recommandation, des jeux de données académiques existants, ainsi que des spécificités liées à la recommandation de contenus de presse. Il a également étudié la dépendance aux données personnelles de diverses plateformes proposant de la recommandation de contenus. Enfin, le PEReN a également proposé différents scénarios permettant de conduire de futures études sur des données réelles, avec pour cible des plateformes de type réseau social ou agrégateur de presse.

¹⁰ Faux *streams*, vrai phénomène : le CNM, avec les professionnels pour lutter contre la fraude (communiqué de presse du CNM et lien vers l'étude : https://cnm.fr/wp-content/uploads/2023/01/CP_CN_M Manipulation-des-streams.pdf)

Doter les administrations d'outils de régulation et d'évaluation

Nos projets en 2022

Amplification des contenus de presse

Sur sollicitation du ministère de la Culture, exploration des effets d'un algorithme de recommandation n'utilisant pas les données personnelles, en particulier concernant les bénéfices pour l'utilisateur quant à son exposition aux contenus préjudiciables, ainsi que les effets pour les éditeurs de presse induits par le changement de visibilité de leurs contenus.

Artificialité des contenus

À la demande de VIGINUM¹¹, conception d'une librairie capable de détecter les photos de profil de faux compte générées artificiellement (*deepfakes*).

Évaluation de la prévalence de bots sur Twitter

Réalisation, en lien avec l'AMF et le ministère de l'Europe et des affaires étrangères, d'un prototype d'outil de détection de comptes automatiques (dits *bots*) reposant sur la stricte analyse des métadonnées des comptes et permettant ainsi de passer à l'échelle et d'avoir un indicateur n'ayant pas recours à des données personnelles.

Présence des vendeurs sur les marketplaces

Appui à la DGE pour l'analyse du fonctionnement algorithmique des *marketplaces*.

Influence des partenaires commerciaux sur les itinéraires de déplacement

Dans le prolongement de premiers développements initiés en 2021 pour l'ART, fourniture de briques techniques facilitant la comparaison des itinéraires proposés par différentes plateformes de mobilité en analysant les biais éventuels sur le calcul des temps de trajets.

Analyse des bannières de cookies

Développement pour les besoins de la CNIL d'un outil permettant de référencer les cookies stockés sur le navigateur d'un utilisateur avant et après son interaction avec des bannières de cookies.

Virilité de groupes de hashtags sur Twitter

Conception d'un tableau de bord opérationnel de suivi de groupes de *hashtags* sur la plateforme Twitter à usage de VIGINUM.

Détection de fausses écoutes (*fake streams*) en ligne

Accompagnement du CNM par la production d'une analyse de l'état de l'art en matière de détection des *fake streams* (création artificielle de fausses écoutes de *streaming*) et du phénomène sur différentes plateformes.

Follow the money des sites gris

Mise à disposition de briques d'analyse des services utilisés par les sites internet impliqués dans la diffusion de contenus problématiques (portant par exemple atteinte à la propriété intellectuelle). L'objectif est d'identifier les acteurs participant le plus au financement et au fonctionnement de ces sites (régies publicitaires et annonceurs, programmes d'affiliation, services de mesure d'audience et de traçabilité, systèmes de paiement en ligne, etc.).

Rapports des plateformes sur la modération des contenus

Accompagnement de l'Arcom dans la conception des rapports de transparence demandés aux plateformes relatifs à la modération des contenus par leurs services, l'étude des algorithmes de modération et la définition du suivi d'indicateurs publiés.

¹¹ Service technique et opérationnel de l'État chargé de la vigilance et de la protection contre les ingérences numériques étrangères

Détection automatique de pratiques commerciales trompeuses

Appui à la DGCCRF dans la détection automatique de pratiques commerciales trompeuses via le développement d'un outil dédié à partir de cas d'usage.

Obligation de portabilité : identification de l'exhaustivité des données personnelles téléchargeables

En lien avec la CNIL, étude comparative entre les données collectées par les plateformes et celles effectivement retournées à l'utilisateur au titre de l'exercice de son droit à la portabilité ou de son droit d'accès.

Consommation de bande passante des services de média à la demande sur différents terminaux

En lien avec la DGE, réalisation d'une étude sur l'adaptation de la qualité du flux vidéo et de sa consommation en bande passante selon le terminal de l'utilisateur pour des plateformes de média à la demande.

Valeur du consentement

Appui à la CNIL pour réaliser une analyse exploratoire de l'impact du parcours web d'un utilisateur sur le niveau des enchères publicitaires menées par les annonceurs pour afficher leur publicité.

Partie 3

Déployer des ressources mutualisées

Placé au service de l'ensemble des administrations en charge de la régulation des plateformes numériques, depuis sa création le PEReN conçoit ses travaux dans une logique de mutualisation. Pour cela, il développe une offre commune de services et de ressources modulables en fonction des besoins de ses partenaires et mobilisables par eux de manière autonome.

Dark Patterns : exploration les interfaces trompeuses

La problématique des *Dark Patterns*, que l'on pourrait caractériser comme « des éléments d'interfaces visant à manipuler ou influencer les utilisateurs en ligne à leur détriment », est de plus en plus présente dans le débat public. Bien que leur définition précise ne fasse pas consensus, les travaux à leur sujet se sont multipliés ces dernières années, tant au niveau académique pour les caractériser précisément, qu'au niveau juridique pour encadrer leur utilisation.

Ainsi, face à l'importance croissante des *Dark Patterns* dans notre environnement numérique, le PEReN a créé et animé un groupe de travail avec les nombreuses administrations publiques concernées par l'étude de ces objets : les ministères de l'Économie et de la culture, l'Autorité de la concurrence, la CNIL et l'Arcom. La mission de ce groupe de travail ? Faire un état des lieux des connaissances scientifiques mais aussi des textes réglementaires sur les *Dark Patterns*, assurer une veille documentaire sur le sujet et réfléchir aux approches techniques permettant de les analyser.

En 2022, ce groupe de travail a permis de faire se rencontrer les différentes visions juridiques des *Dark Patterns* (droit de la consommation, de la concurrence, données personnelles...) mais aussi de les confronter aux définitions académiques. Un travail de synthèse a été réalisé pour permettre aux administrations de partager une taxonomie et des connaissances juridiques et scientifiques communes pour l'étude de ces objets. Au sein de cette instance d'échange, le PEReN a pu apporter son expertise technique pour la conception d'un prototype permettant d'identifier et d'analyser automatiquement certains types de *Dark Patterns*. La conception et l'amélioration de cette preuve de concept se poursuit en 2023 et constituera une nouvelle brique logicielle réutilisable par toutes les administrations concernées selon leur périmètre d'intervention.

Le traitement automatique du langage au service de l'analyse de corpus textuels volumineux

Les autorités indépendantes et les administrations peuvent avoir à exploiter de grands volumes de corpus de données textuelles dans le cadre de leurs missions de régulation des plateformes numériques (par exemple, pour l'analyse d'avis en ligne ou d'efficacité de systèmes de modération). La mise en place de processus de traitement automatique du langage naturel (*Natural Language Processing – NLP*) pour traiter de telles données nécessite des ressources techniques dont la plupart des administrations ne disposent pas.

Dans ce cadre et pour répondre aux différents besoins de ses partenaires comme des siens, le PEReN a développé une interface de programmation applicative (*Application Programming Interface – API web*). Fonctionnellement, cette brique logicielle de base vise donc à fournir des services de traitement automatisé du langage à l'état de l'art et faciles à mobiliser par des autorités ou administrations ayant des équipes techniques de taille réduite et n'ayant pas les capacités de concevoir, déployer et maintenir ce type d'application.

Sur la base de ce socle applicatif, des services génériques de base, utiles à de nombreux projets, ont été réalisés en 2022. Ils portent, d'une part, sur la classification de textes et plus particulièrement la classification binaire utilisable dans différents domaines tels que la détection de *fake news* ou de contenus haineux, et d'autre part, sur la comparaison d'unités textuelles de niveau phrase ou paragraphe.

Par nature, ces outils applicatifs ont vocation à évoluer en fonction des besoins des différents projets conduits par le PEReN. En 2023, ces développements seront consolidés techniquement.

Regrouper, classer : merci l'API *Clustering* !

Des projets confiés par nos administrations ou autorités de régulation partenaires peuvent souvent nécessiter d'une manière ou d'une autre de partitionner de vastes ensembles de données sans pour autant que soient disponibles des étiquettes de classification explicites. Reposant sur une méthode d'analyse statistique et algorithmique, les techniques de *clustering* sont une solution à ce problème en permettant d'organiser des données brutes en groupes homogènes (*clusters*).

Dans sa démarche de mutualisation d'outils entre ses différents projets, le PEReN a conçu et déployé une brique logicielle de base permettant d'accomplir des tâches de *clustering*, ou partitionnement de données, de manière simplifiée, par de simples appels à une interface applicative en ligne (API web).

En 2022, des services de *clustering* mettant en œuvre différents algorithmes ont été réalisés :

- *Clustering* à partir de textes : l'utilisateur de l'API a juste à téléverser des fichiers texte dont il souhaite effectuer le classement automatique en groupes cohérents.
- *Clustering* à partir de matrices : réservé à un utilisateur plus averti, ce service permet de téléverser une matrice construite à partir de textes (*embeddings*) et de regrouper les lignes de cette matrice, chacune d'elles représentant le contenu d'un texte.

Pour s'ajuster toujours au mieux aux différents besoins du PEReN et de ses partenaires, la consolidation technique de ce service applicatif se poursuivra en 2023.

Identification des données personnelles transmises sans consentement par les applications mobiles

Dans le cadre de ses missions visant à garantir le respect du Règlement Général sur la Protection des Données (RGPD) et de la loi Informatique et Libertés du 6 janvier 1978, la CNIL est amenée à identifier et contrôler les services en ligne qui effectueraient des opérations de lecture et écriture sur le terminal d'un utilisateur sans son consentement, notamment la lecture d'identifiant publicitaire.

Aux fins d'étendre et simplifier ses contrôles sur les applications sur *smartphone*, la CNIL a sollicité l'appui technique du PEReN. L'objectif : développer un applicatif facilitant la procédure de constat et rendant possible une analyse semi-automatique des applications.

Dans cette optique, le PEReN a conçu un outil permettant d'identifier les accès, sans consentement, à l'identifiant unique d'un utilisateur sur *smartphone* (le « *Google Advertising ID* » ou GAID). À destination des agents de la CNIL, cet outil a également vocation à faciliter les procédures en enregistrant l'état et les communications du *smartphone* utilisé pour le contrôle afin de produire plus rapidement les constats.

Conjointement à d'autres projets, le développement de cet outil d'étude des données personnelles transmises par les applications mobiles, dont le code source a été ouvert¹², a contribué à renforcer les compétences du PEReN en matière d'instrumentation des *smartphones*. En 2023, les développements dans ce domaine auront vocation à construire une véritable boîte à outil pour l'exploration des usages des applications sur *smartphones*.

¹² Code source du projet ouvert sur la plateforme code.peren.fr, sous la référence « constapp »

Déployer des ressources mutualisées

Nos projets en 2022

Partitionnement automatisé de données (*clustering*)

Conception et déploiement d'une interface de programmation applicative en ligne (API web) permettant aux partenaires du PEReN et à son équipe d'accomplir des tâches de *clustering*, ou partitionnement de données, de manière simplifiée, c'est à dire sans nécessiter la mise en œuvre d'un entraînement coûteux et basé sur des techniques d'étiquetage.

Analyse de corpus textuels volumineux par traitement automatisé du langage

Dans une logique de mutualisation, mise à disposition de tous les partenaires du PEReN, et de son équipe, d'une interface de programmation applicative en ligne (API web) offrant des fonctionnalités d'analyse de texte (détection de *ngrams*, d'entités nommées ou de similarités entre unités textuelles).

Collecte automatisée de données publiquement accessibles (*scraping*)

Dans une logique de mutualisation, mise à disposition de tous les partenaires ayant des besoins de *scraping*, et justifiant des compétences juridiques requises, d'une interface de programmation en ligne (API web) de collecte automatisée de données publiquement accessibles.

Exploration des interfaces trompeuses ou *Dark Patterns*

Animation d'un groupe de travail destiné à faire un état des lieux des connaissances scientifiques mais aussi des textes réglementaires sur les *Dark Patterns*, assurer une veille documentaire sur le sujet et réfléchir aux approches techniques permettant de les analyser.

Projet *Privacy Sandbox* de Google

Animation d'un groupe de travail réunissant toutes les administrations intéressées (ministères de la Culture et de l'Économie, Autorité de la concurrence, CNIL, Arcom) et dédié à l'analyse technique mutualisée de la future *Privacy Sandbox* de Google et à une veille sur le sujet.

Qualité de service des messageries OTT

A la demande de l'Arcep, conception d'une interface applicative permettant d'automatiser des tests qualité sur des services de messageries instantanées OTT (*Over The Top* ou par contournement) sous différentes conditions d'utilisation réalistes de l'appareil (géolocalisation, batterie, qualité du réseau...) (code *open source* de l'outil : <https://code.peren.fr/peren/2022-qualit-de-service-des-messageries-ott/qualott>).

Applications mobiles : identification des données personnelles transmises

En appui à la CNIL, développement d'un outil facilitant l'identification des accès par des applications mobiles à des données personnelles, avec ou sans consentement préalable.

Suivi automatique des rapports environnementaux des plateformes

Sur demande de la DGE, mise en œuvre d'un outil de collecte automatique des rapports environnementaux dont la publication par les opérateurs de plateforme est une obligation réglementaire, en vue d'en exploiter le contenu pour réaliser des analyses comparatives et suivre leur évolution (code *open source* de l'outil : <https://code.peren.fr/peren/2022-analyse-de-rapports>).

Cadre juridique du *scraping*, échanges sur les pratiques

Animation d'un groupe d'échanges autour des pratiques de collecte automatisées de données (*scraping*) mises en œuvre par les différentes organisations publiques et du suivi des évolutions techniques et juridiques en la matière afin de suggérer de potentielles évolutions du cadre législatif ou des pratiques.

Partie 4

Travailler à l'état de l'art et diffuser les savoirs

Face aux évolutions technologiques constantes dans le domaine de l'intelligence artificielle et de l'analyse des données, deux fondamentaux s'imposent au PEReN : travailler à l'état de l'art pour assurer à ses partenaires une expertise technique toujours à la pointe de la recherche et, agrégateur d'expertises par nature, contribuer à la diffusion des savoirs au sein de l'écosystème de la régulation du numérique.

Exploration scientifique au service de la régulation des plateformes numériques de travail

En 2021, le PEReN avait réalisé différents travaux de recherche portant sur les questions d'anonymisation des données et d'auditabilité des algorithmes, certains en collaboration avec la CNIL ou avec Inria. Ayant permis de capitaliser une grande expertise, ces travaux ont pu paraître, à certains égards, peu accessibles auprès de certains partenaires ou acteurs de la société civile. Pour cette raison, en 2022, le PEReN a entrepris la promotion de cette expertise à la fois auprès de chercheurs plus proches des sciences humaines et sociales que de l'informatique, ainsi qu'auprès de l'Autorité des relations sociales des plateformes d'emploi (ARPE) et de certains partenaires sociaux qui lui sont liés.

En pratique, cette promotion s'est matérialisée par :

- la participation du PEReN au colloque pluridisciplinaire, NUTRA 2022¹³, dont les actes à paraître prochainement intégreront l'article scientifique sur les travaux menés ;
- la présentation de travaux du PEReN au colloque du groupe de travail « vie privée » du CNRS, APVP 2022¹⁴ (Atelier de Protection de la Vie Privée) ;
- une présentation à visée pédagogique des travaux et enjeux à des représentants de l'ARPE et de partenaires sociaux.

Ces temps de présentation et d'échanges ont permis de vulgariser les notions générales d'algorithmes, de transparence et différents termes techniques trop souvent mal compris dans le domaine de l'intelligence artificielle. Ils ont également permis de faire la pédagogie, d'une part, de travaux plus proches de l'état de l'art sur l'anonymisation et l'auditabilité, en

¹³ La plateforme du travail : nouvelles formes d'emploi, nouvelles formes d'organisation ? (<https://nutra2022.sciencesconf.org/resource/page/id/2>)

¹⁴ « Le PEReN intervient à l'Atelier sur la Protection de la Vie Privée (APVP) » (Actualité en ligne : https://www.peren.gouv.fr/actualites/2022-06-29_participation_apvp/)

illustrant systématiquement par des exemples concrets, et, d'autre part des enjeux portant sur la régulation.

Cette démarche de vulgarisation des enjeux techniques liés à la régulation numérique et des travaux du PEReN a naturellement vocation à se consolider au travers des différentes actions de communication et de médiation entreprises sur le champ particulier de la régulation des plateformes numériques.

AMP de Google : les dessous d'une technologie d'optimisation d'affichage d'une page web

Pour améliorer la navigation sur mobile, Google a déployé en 2015 une technologie multifacette : AMP pour *Accelerated Mobile Pages*. Cette technologie désigne un ensemble d'outils destinés à accélérer l'affichage de pages web sur mobile. Elle consiste en une série de règles destinées à proscrire l'utilisation de fonctionnalités lentes par les développeurs et matérialisées par un langage proche du HTML. À cela s'ajoutent deux mécanismes activés par défaut sur la plupart des navigateurs lors de la consultation d'une page AMP depuis un moteur de recherche : la mise en cache ou mémorisation et le pré-rendu ou anticipation de l'apparence de cette page.

Prometteuse, AMP a rapidement été adoptée par de nombreux éditeurs de contenu conscients des usages grandissants d'Internet sur mobile et attirés par des avantages procurés en matière de référencement. Las, ces mêmes éditeurs ont dû s'adapter aux multiples contraintes et évolutions techniques imposées par cette technologie. De l'organisation des contenus aux fonctionnalités de monétisation et de suivi des utilisateurs, la complexité de mise en œuvre a conduit beaucoup d'éditeurs à reconsidérer leur usage d'AMP : déjà abandonnée par certains, d'autres envisagent une sortie imminente. La fin, en 2021, de l'avantage compétitif accordé par Google aux pages AMP dans les résultats de son moteur de recherche semble avoir conforté le mouvement enclenché.

Pour explorer l'ensemble de la dynamique à l'œuvre, le PEReN a conduit en 2022 une étude technique s'appuyant sur plusieurs entretiens réalisés avec des éditeurs de contenu et des tests en conditions réelles dont le code a été mis en ligne en open-source¹⁵. Il a communiqué les résultats de ce travail dans le n°5 de la série « Éclairage sur... »¹⁶ (octobre 2022) qui propose une plongée dans ce dispositif de navigation contesté pour en comprendre le fonctionnement, les avantages et désavantages techniques et économiques, ainsi que les enjeux contemporains ou futurs.

¹⁵ Code source pour reproduire les expériences conduites dans l'étude (<https://code.peren.fr/peren/2022-AMP/analyse-de-l-amp-de-google>)

¹⁶ Éclairage sur...n°5 - AMP de Google : du dopage au sevrage ? (https://www.peren.gouv.fr/rapports/2022-10-17%20-%20Eclairage-sur-AMP_FR.pdf)

« Possibilités technologiques, limites juridiques : les algorithmes face à la régulation » (Conférences et Hackathon 2022)

Le 30 novembre 2022, sous le haut patronage du Ministre délégué chargé de la Transition numérique et des Télécommunications, le PEReN a organisé avec l'appui de la DGE, et ses partenaires Inria, le CNNum (Conseil national du numérique) et le département d'économie de Sciences Po Paris, son premier événement annuel : « Possibilités technologiques, limites juridiques : les algorithmes face à la régulation ». Rassemblant à Bercy plus de 180 participants, cette journée originale a allié théorie et pratique autour de conférences et du hackathon « Segmente-moi si tu peux ! ».

Cette rencontre s'est inscrite dans une période charnière pour la réglementation européenne et nationale de l'espace numérique. Son ambition ? Apporter une vision pluridisciplinaire des enjeux pour affiner les choix de régulation et aborder le plus concrètement possible les questions de mise en œuvre des nouvelles obligations. Ainsi, juristes, scientifiques, entreprises et administrations, entre autres, ont exploré ensemble les possibilités et contraintes de traduire techniquement les concepts liés aux notions de causalité, d'intentionnalité, d'accord de volonté et de responsabilité, au bénéfice d'une régulation toujours plus pertinente des plateformes numériques. Ils ont également questionné le bon équilibre en matière de régulation au regard notamment des droits fondamentaux et de l'innovation.

Côté hackathon, 24 codeurs se sont mesurés au challenge concocté par le PEReN autour d'un cas fictif, mais réaliste, de plateforme de livraison de plats cuisinés. Les prix de livraison étant définis par différents algorithmes, les concurrents ont eu à les identifier pour répartir ensuite les restaurants en fonction de la méthode de tarification pratiquée.

Entrant en résonance avec les problématiques des conférences, ce concours de code a illustré le procédé d'analyse externe d'algorithme exploré par le PEReN. Son concept ? Prototyper un outil concret fondé sur une méthode d'exploration dite en « boîte noire », c'est-à-dire non intrusive et sans accès au code source, au modèle ou aux données qui composent le système d'intelligence artificielle. Empruntant la démarche scientifique du *testing*, l'approche du PEReN repose sur l'analyse statistique de réponses d'algorithmes sur des entrées soigneusement choisies pour en déduire des grands principes de fonctionnement ou des biais locaux.

Dans ce champ exploratoire, après avoir validé, en 2021, à partir de données artificiellement élaborées, le socle méthodologique et technique de son approche pour identifier de grands paramètres algorithmiques d'un algorithme de tarification, le PEReN a testé son modèle d'analyse in vivo.

Travailler à l'état de l'art et diffuser les savoirs

Nos projets en 2022

Appui sur la recherche

Rencontres académiques autour des applications techniques du DSA

Organisation d'entretiens avec des représentants du monde de la recherche pour un état des lieux académique en matière d'audit algorithmique au regard des besoins techniques des régulateurs. Ces rencontres ont pour objectif de permettre la rédaction de lignes directrices pour les applications techniques du DSA.

Faisabilité des approches de l'*Election Integrity Partnership*

Initiative du *Stanford Internet Observatory* et de l'Université de Washington, l'*Election Integrity Partnership* vise à fournir aux académiques et à la société civile des outils et des méthodes d'évaluation de la prévalence de contenus nuisibles au processus démocratiques. Le PEReN a fourni un support technique à des chercheurs du centre GEODE pour évaluer la possibilité d'appliquer les mêmes méthodes au contexte français.

Détection et caractérisation de contenus viraux

Prolongement du projet lancé en 2021 à la demande du ministère de la Culture. En application de travaux académiques, il vise à développer une méthodologie de détection des contenus viraux susceptibles d'être problématiques en s'appuyant sur leurs propriétés de propagation et non sur une analyse directe du contenu ou des personnes impliquées. Focus 2022 : quantifier la notion de viralité à partir d'indicateurs statistiques en se concentrant sur les contenus à forte visibilité.

Anonymisation des données des travailleurs des plateformes

Initialisation des travaux pour l'élaboration d'un outil destiné aux travailleurs des plateformes leur permettant de visualiser les données qu'ils peuvent retrouver sur leur propre compte propre (par exercice du droit à la portabilité du RGPD) et de pouvoir les communiquer à des tiers agrégateurs après anonymisation.

Comparaison des algorithmes d'autocomplétion

Sur la base d'une bibliographie académique, exploration du fonctionnement d'algorithmes d'autocomplétion (assistance à la saisie par suggestion de mots, phrases ou expressions) de plateformes numériques.

Création de *personas* pour l'API de *scraping*

Création de *personas* (profils web d'utilisateurs fictifs définis par un ensemble de cookies) afin d'étudier l'impact d'algorithmes ou de traitements différenciés sur le web en fonction de segments de profils utilisateurs. Ces outils constituent un développement d'intérêt académique qui a été présenté au colloque APVP (Atelier de Protection de la Vie Privée) en 2023.

Étude de la faisabilité de méthodes pour une analyse de contenus multimodaux

Exploration de techniques d'apprentissage faiblement supervisées permettant de traiter conjointement images et textes associés sur les plateformes numériques. L'étude préalable de l'état de l'art a permis notamment d'identifier les architectures à choisir.

Diffusion des savoirs

Séminaires académiques

Animation de rencontres trimestrielles ayant pour objectif de favoriser les échanges entre administrations en charge de la régulation des plateformes numériques et le monde de la recherche. Lors de ces rencontres, un chercheur ou un expert est invité à présenter ses travaux sur des thématiques en lien avec les activités du PEReN.

Collection « Éclairages sur... »

Publications trimestrielles dans le cadre la collection « Éclairage sur... » qui propose, dans un esprit de vulgarisation, des éléments d'analyse techniques sur des thèmes liés à la régulation des plateformes numériques (numéros parus disponibles ici : <https://www.peren.gouv.fr/publications/>).

Ouverture de codes source

Publication en *open source*, dans un espace dédié (<https://code.peren.fr/peren/>), des codes de services applicatifs et outils mis en production.

Organisation d'un événement annuel grand public

Sous le haut patronage du Ministre délégué chargé de la Transition numérique et des Télécommunications, avec l'appui de la DGE et en partenariat avec Inria, le CNNum et Sciences Po Paris, organisation d'un événement grand public alliant conférences et hackathon pour mobiliser l'intelligence collective autour des enjeux liés à la régulation numérique.

Décryptage d'AMP, technologie Google d'optimisation d'affichage d'une page web

Étude visant à comprendre le fonctionnement détaillé de la technologie AMP de Google, d'analyser la pertinence des différents arguments avancés par Google (impact de la durée de chargement, mise en cache sur d'autres serveurs...) et d'évaluer de potentielles alternatives à cette technologie (étude disponible ici : https://www.peren.gouv.fr/actualites/2022-10-17_eclairage_sur_amp/).

Traitement automatique des rapports de transparence

Réalisation d'un outil de visualisation des rapports de transparence que les opérateurs de plateforme ont l'obligation de publier et de mise en forme dans un format facilement exploitable manuellement ou automatiquement.

Impact des algorithmes sur la requalification des travailleurs

Identification et analyse des jurisprudences et textes légaux (en vigueur et en discussion) portant sur les algorithmes des plateformes numériques de travail afin de mieux comprendre les exigences légales et les moyens utilisés pour caractériser ces algorithmes et leur impact. Promotion de cette expertise à la fois auprès de chercheurs et autorités de régulation.

Partie 5

Nouveaux pouvoirs et montée en puissance

Pionnière et innovante, la création du PEReN a eu pour but de doter l'État d'un pôle d'excellence en matière de science des algorithmes et des données. Cela a contribué à consolider le dialogue des pouvoirs publics avec les plateformes numériques, de renforcer l'action de la France en matière de régulation et de mutualiser les connaissances entre services de l'État et autorités. En 2022, le cadre d'intervention du PEReN a franchi un nouveau cap : le législateur l'a doté des pouvoirs d'expérimenter et de conduire des travaux de recherche publique.

En 2022 le cadre d'intervention du PEReN évolue : l'expérimentation est permise !

Afin de faire face à l'opacité des systèmes algorithmiques opérés par les plateformes numériques et à la dépendance vis-à-vis de ces acteurs pour obtenir des informations permettant de mieux comprendre leur fonctionnement et de définir une régulation adaptée, le législateur a doté le PEReN du pouvoir d'expérimentation¹⁷.

Une expérimentation consiste à la phase de conception « à blanc » d'un outil, pendant laquelle les plateformes ont une obligation de coopération et ne peuvent bloquer le PEReN. Dans cette phase, tout est fait pour que les tests conduits ne donnent lieu à aucune perturbation pour les utilisateurs ni pour les plateformes. Les données récoltées pendant cette phase sont ensuite détruites, et si des services de l'État ou une autorité indépendante dans leurs missions de régulation des plateformes numériques peuvent être destinataires de l'outil qui a été conçu pendant l'expérimentation, aucune donnée collectée n'est transmise. La conduite de ces activités d'expérimentation a ainsi été encadrée par le législateur, tout comme les modalités de collecte de données que le PEReN pourrait mettre en œuvre dans ce cadre¹⁸.

Avec cette nouvelle modalité, le PEReN peut désormais intervenir de trois manières différentes : à la demande des autorités indépendantes ou des administrations dans le cadre

¹⁷ Article 36 de la loi n°2021-1382 du 25 octobre 2021 relative à la régulation et à la protection de l'accès aux œuvres culturelles à l'ère numérique (<https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000044245615>)

¹⁸ Décret d'application n° 2022-603 du 21 avril 2022 fixant la liste des autorités administratives et publiques indépendantes pouvant recourir à l'appui du Pôle d'expertise de la régulation numérique et relatif aux méthodes de collecte de données mises en œuvre par ce service dans le cadre de ses activités d'expérimentation (<https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000045638527>)

d'une convention, au travers d'expérimentations visant à tester la faisabilité d'outils de régulation ou bien en conduisant des travaux de recherche publique.

Dès l'adoption du décret d'application en avril 2022, le PEReN s'est emparé de cette nouvelle compétence d'expérimentation en finalisant le cadre juridique permettant la collecte des données ainsi qu'en mettant en place l'infrastructure technique spécifique à cette nouvelle mission. En juillet 2022, il a lancé ses 5 premiers projets expérimentaux au sujet desquels il a remis en mai 2023, au Parlement et à la CNIL, un rapport annuel d'évaluation, conformément à ses obligations réglementaires. Focus sur 2 d'entre eux !

Analyse du type de réseau routier proposé aux automobilistes par les calculateurs d'itinéraires

La loi Résilience et Climat prévoit de nouvelles obligations pour les calculateurs d'itinéraires et notamment quant à l'usage qu'ils font du réseau routier principal et secondaire.

L'objectif de cette expérimentation a été triple :

- développer les briques technologiques nécessaires à la collecte des itinéraires proposés par des calculateurs d'itinéraires au travers de leurs différentes interfaces (web, API pour les développeurs et applications mobiles) ;
- élaborer des méthodes automatiques de rapprochement de ces itinéraires, dans des formats de données propriétaires, avec les bases de données géographiques nationales ou libres du réseau routier afin de permettre la catégorisation automatique des axes empruntés en tant que partie du réseau routier principal ou secondaire ;
- orchestrer ces différents composants afin de permettre une industrialisation de ces analyses en vue de permettre aux services de l'État d'analyser le respect ou non de leurs obligations par les calculateurs d'itinéraires.

Cette expérimentation a permis de valider techniquement l'intégralité de ces briques logicielles afin de réaliser des analyses des itinéraires proposés par les calculateurs d'itinéraire. Ces travaux vont permettre d'accélérer la mise à disposition de ces outils aux administrations concernées.

« Bulle de filtre » ou enfermement informationnel : développement d'outils d'exploration de ce phénomène algorithmique

De nombreuses plateformes, notamment de réseaux sociaux, s'appuient sur des algorithmes de recommandation afin de proposer du contenu personnalisé à leurs utilisateurs. Exploitant l'historique d'interactions, elles parviennent à déterminer, parfois avec grande précision, leurs centres d'intérêt. Dans certains cas, une trop forte personnalisation de ces recommandations peut conduire à une perte importante dans la diversité des contenus proposés par la plateforme. Ainsi, dans le cas d'un réseau social de partage de vidéos, l'utilisateur peut se voir recommander quasi exclusivement des vidéos portant sur un sujet spécifique, on parle alors de « bulles de filtre ». Déjà mises en évidence par plusieurs équipes de recherche¹⁹, les bulles de filtre ne sont pas problématiques en soi, mais dans certains cas, elles peuvent se révéler nocives, induisant des biais de confirmation chez certains utilisateurs sur des sujets sensibles.

¹⁹ À citer notamment : *Degenerate Feedback Loops in Recommender Systems* (<https://arxiv.org/pdf/1902.10730v3.pdf>)

En 2022, le PEReN a développé un prototype d'outil d'identification et d'analyse des bulles de filtre par traitement automatisé du langage (TAL). Ce dispositif s'appuie sur les mots clés issus de la description d'une vidéo sur un réseau social afin d'en identifier la thématique. Des simulations automatisées de parcours utilisateurs sur la plateforme permettent ensuite de corréliser l'augmentation de la prévalence de certaines thématiques dans les recommandations à l'historique de visionnage, ce qui permet *in fine* de détecter d'éventuelles bulles de filtre.

L'expérimentation a permis de mettre en évidence les limites d'une telle approche n'utilisant que la description des vidéos, sans aucune analyse de leur contenu : sur un certain nombre de plateformes de réseau social, la description n'est qu'un appui au référencement de la vidéo mais ne constitue aucunement un résumé de celle-ci.

Aussi, en 2023, le projet se poursuit comme projet de recherche publique, avec pour objectif de quantifier plus précisément le phénomène de bulle de filtre en exploitant directement le contenu des vidéos. Ces travaux, synthétisés dans une courte vidéo, ont été présentés à l'invitation de la Commission européenne à l'événement de lancement du Centre européen pour la transparence algorithmique de Séville (ECAT)²⁰.

2022 : année de consolidation

Si 2021 a été l'année de l'installation du PEReN, marquée par une forte accélération dans le déploiement des effectifs et des ressources matérielles, 2022 a résolument été l'année de la consolidation.

Engagé dans une dynamique de croissance rapide, le PEReN a su s'installer, en un peu plus de deux ans, comme un partenaire de référence au sein de l'écosystème de la régulation des plateformes numériques. Témoin de l'attractivité de son offre de service et de l'adéquation de son expertise aux besoins de ses partenaires, services de l'État comme autorités administratives indépendantes, son programme de travail 2022 a compté une soixantaine de projets, soit trois fois plus qu'en 2021.

Comment le PEReN définit-il son programme de travail ?

Pour élaborer sa feuille de route annuelle, le PEReN associe étroitement les administrations partenaires à ses réflexions, dans le cadre d'échanges bilatéraux avec chacune pour déterminer ses besoins propres, et de réunions multilatérales pour présenter collectivement la nécessaire priorisation des projets. La sélection finale des projets qui seront réalisés par le PEReN l'année suivante s'opère au regard de la proportionnalité entre la charge de travail induite et la valeur pour le partenaire, de sa complexité technique, de la faisabilité dans le cadre juridique de chaque partenaire, et lorsque cela est possible, des possibilités de partage entre administrations ou de mise à disposition du public.

Conformément à son décret de création, le PEReN soumet chaque année sa feuille de route à l'approbation de ses trois ministres de tutelle (Économie, Communication, Numérique).

²⁰ Vidéo de démonstration du dispositif expérimental (<https://www.peren.gouv.fr/publications/>)

Problématique émergente : répondre à une demande sans cesse croissante

Le PEReN fonde sa spécificité en premier lieu sur la constitution d'une équipe de *data scientists* et développeurs mobilisant des compétences et savoirs pointus et rares en matière de sciences des données (*machine learning* / intelligence artificielle, programmation informatique, algorithmique) et des expériences de recherche académique. Ensemble, ils donnent au PEReN sa marque de fabrique : rassembler au sein d'un service de l'État tout à la fois des capacités à concevoir, développer et analyser.

Initialement doté d'un plafond d'emploi de 20 ETP (Équivalent-Temps-Plein), le Pôle a atteint son effectif cible fin 2022, tandis que le volume annuel d'activité reste croissant. Cette année a ainsi marqué le début d'une nécessaire priorisation des projets en concertation avec les partenaires, et notamment en conservant un équilibre entre les différents services de l'État et administrations, et en favorisant les projets nécessitant le recours aux technologies les plus pointues ou offrant le maximum de potentiel de mutualisation. La dynamique observée a montré également la nécessité de préserver des ressources disponibles pour être en mesure de répondre en cours d'année à des besoins non anticipés dans le programme annuel mais néanmoins prioritaires.

En 2023, le plafond d'emploi a été porté à 25 ETP pour demander au PEReN de répondre aux enjeux nouveaux dans le champ des modèles d'IA générative et des giga-modèles de langage (LLM). Le PEReN a pour ambition d'atteindre cet objectif dès l'automne 2023, avec des recrutements permettant le développement d'une expertise renforcée en ces domaines, tout en structurant son organisation avec la création de rôles de référents en Recherche et éthique, *Data science* et génie logiciel. Ces évolutions doivent permettre de mieux faire face à une demande d'extension du champ d'expertise tout en se préparant à la croissance attendue des ressources qui seront consacrées à la mise en œuvre des nouvelles régulations numériques et mutualisées entre régulateurs nationaux et européens dans un souci d'efficacité maximale.

Le PEReN travaille depuis trois ans à la mise en place des outils de régulation numérique dans des champs variés. Son approche modulaire des projets qui lui sont confiés s'appuie sur le développement de différentes briques logicielles combinables et reconfigurables selon les besoins. Éprouvé, ce modèle permet d'absorber un grand volume de projets et de produire tout autant de résultats, tout en mutualisant le plus possible les outils développés. En l'état, ce processus reste adapté dès lors que suffisamment de partenaires sont capables d'anticiper leurs besoins, ce qui permet au PEReN de lisser la charge de travail sur l'ensemble de l'année et de répondre à un nombre important de commandes avec des effectifs comparativement réduits.

Des infrastructures informatiques sécurisées et homologuées

Capitaliser au sein de l'État un ensemble de compétences académiques et techniques est un des objectifs qui avait été assignés au PEReN, qui a développé une organisation et des ressources lui permettant de conduire en interne l'ensemble des étapes d'un projet, sans recours à de la prestation extérieure ni à des logiciels propriétaires.

Sur le plan technique, cette capacité a été rendue possible par le développement d'une architecture en propre pour, d'une part, maîtriser intégralement les choix technologiques et

les données susceptibles d'être collectées et stockées, et d'autre part, inscrire l'exploitation des outils dans la durée et permettre leur utilisation par les partenaires sans surcoût d'usage.

Ce choix se traduit donc par :

- la possession de serveurs pleinement opérationnels permettant d'assurer les besoins internes du service mais également l'hébergement des interfaces de programmation applicative (API) mises à disposition des organisations publiques partenaires ;
- la présence au sein de l'équipe de deux membres en charge de l'administration du système d'information et de sa sécurité.

En 2022, le PEReN a poursuivi la construction de l'ensemble de cette infrastructure personnalisée avec en particulier la mise en place :

- d'une forge logicielle en propre (système de gestion et de maintenance collaborative entourant un développement logiciel) et de bases de données ;
- de bases de données à accès restreints et journalisés pour les projets relevant du cadre d'expérimentation ;
- d'outils de supervision.

Pour s'assurer de la bonne mise en œuvre de la politique de sécurité des systèmes d'information de l'État (PSSIE), le PEReN s'est engagé depuis 2021 dans une démarche d'homologations successives de son infrastructure et de ses services applicatifs. Après l'architecture de son système d'information (homologation sans réserve pour une période de 3 ans), c'est son interface de programmation applicative (API) pour la collecte automatisée de données qui a été homologuée en 2022. S'en sont suivis les audits des API de clustering (partitionnement des données) et de NLP (analyse de corpus textuels volumineux) en vue d'une homologation courant 2023.

Le PEReN a par ailleurs mis en place pour l'ensemble de ses agents une sécurisation des connexions à ses plateformes collaboratives de travail via des certificats d'authentification sur clés physiques, ainsi qu'un dispositif d'habilitation par projets via la signature authentifiée des membres des équipes projets et du directeur du PEReN.

Nos expérimentations en 2022

La loi du 25 octobre 2021, relative à la régulation et à la protection de l'accès aux œuvres culturelles à l'ère numérique, a doté le PReN du pouvoir d'expérimentation. En 2022, première année de mise en œuvre de cette nouvelle compétence, cinq projets ont été menés.

Suivi des prix et délais de livraison pratiqués par les places de marché

Étude des possibilités de suivi historisé des évolutions de prix et délais de livraison sur les sites de e-commerce afin de vérifier que les informations disponibles publiquement et pouvant être collectées sont suffisantes pour permettre des analyses par les services compétents ou, dans un second temps, l'élaboration d'algorithmes de détection de mauvaises pratiques. L'expérimentation a permis de valider la mise au point d'un outil de suivi qui pourra être utilisé par des administrations intéressées.

Création de profils de test

Génération de profils d'utilisateurs types, définis par un historique de navigation (*cookies*), dans le but d'étudier l'impact de la personnalisation sur les résultats fournis par les algorithmes ou les interfaces d'une plateforme. Cette expérimentation a confirmé la possibilité de créer des profils personnalisés exploitables pour l'étude du fonctionnement d'algorithmes de recommandation. Toutefois, la mutualisation de ces profils types reste limitée en raison du code informatique requis très dépendant de la plateforme potentiellement étudiée. Sans rechercher un outil universel, des outils spécifiques reposant sur cette expérimentation pourront être développés au cas par cas.

Convergence algorithmique des contenus

Étude de la faisabilité d'un outil capable de réaliser automatiquement un parcours utilisateur sur une plateforme en ligne de vidéos à la demande selon des préférences déterminées, sans stockage des vidéos ni analyse de leur contenu mais uniquement en utilisant la description textuelle qui en est faite. Un tel outil permettrait d'étudier les algorithmes de recommandation, notamment s'agissant des bulles de filtre (enfermement informationnel). Si l'expérimentation a permis d'élaborer un outil de collecte sur applications mobiles, elle a confirmé que les seules descriptions des vidéos n'étaient pas suffisantes pour une analyse du phénomène de bulle de filtre. L'analyse du contenu de la vidéo et de la bande son d'accompagnement sont apparues nécessaires.

Audit algorithmique en boîte noire

Développement de méthodes d'évaluation des principaux paramètres algorithmiques d'une plateforme, en testant directement son interface publique, c'est-à-dire sans nécessiter sa coopération active. Cette expérimentation a permis de développer des stratégies d'optimisation pour adapter automatiquement les échantillons demandés et ainsi minimiser le volume de requêtes nécessaires. Elle a ainsi permis de développer des briques logicielles d'audit algorithmique pouvant s'adapter à des stratégies de tarification dynamique ou *surge pricing* (tarification incitative). L'évaluation des principaux paramètres algorithmiques en boîte noire pourrait ainsi permettre de confronter les obligations de transparence consacrées par le règlement européen P2B à l'expérience.

Analyse du type de réseau routier proposé aux automobilistes par les calculateurs d'itinéraires

Aux fins d'évaluation de mise en œuvre des obligations de la loi Climat et Résilience, développement de briques logicielles nécessaires à la collecte automatique des itinéraires proposés par des calculateurs d'itinéraires, au rapprochement de ces itinéraires avec les bases de données géographiques nationales du réseau routier et à la catégorisation automatique des axes empruntés en tant que partie du réseau routier principal ou secondaire. Le dernier objectif de cette expérimentation, à savoir le développement de briques d'analyse de la conformité des itinéraires proposés aux obligations des plateformes, a été limité au vu du faible volume de données disponibles.

Pôle d'expertise de la régulation numérique – PEReN

120 rue de Bercy – 75012 Paris | www.peren.gouv.fr | contact.peren@finances.gouv.fr

Directeur de la publication : Nicolas Deffieux, directeur du PEReN

Conception, rédaction et réalisation : PEReN

Juillet 2023

N° ISSN : en cours

Crédits : ©everythingpossible - stock.adobe.com