

COMMUNIQUE DE PRESSE

Paris, le 28 novembre 2022.

Permettre à la science des données de franchir un cap décisif au service de l'intérêt général : l'apprentissage fédéré, un immense défi scientifique

Alors qu'une plus grande disponibilité de gros volumes de données permettrait des avancées significatives dans de nombreux domaines, notamment en santé (prévention, l'assistance au diagnostic, au pronostic ou à la décision, épidémiologie, pharmacologie...), les questionnements sont multiples. Comment concilier cet enjeu de disponibilité accrue avec le nécessaire respect de la vie privée des patients ou des assurés sociaux ? Comment éviter les coûts et les risques liés à la centralisation et à la transmission de gros volumes de données ? Comment contribuer à l'acceptabilité de l'intelligence artificielle en promouvant un paradigme où les propriétaires de données gardent le contrôle de leurs données et peuvent mieux protéger leurs informations sensibles ? C'est tout l'enjeu du Défi FedMalin¹ qui vient d'être lancé par Inria avec le soutien du groupe La Poste. Cette initiative commune s'inscrit dans le cadre du [partenariat de mécénat](#) noué en 2021 avec la Fondation Inria pour accélérer l'innovation numérique responsable au service de la société.

Un enjeu majeur pour la recherche : améliorer l'accès aux données, véritable « carburant » de l'apprentissage automatique

L'apprentissage automatique² vise à donner aux ordinateurs la capacité d'apprendre à partir de jeux de données, c'est-à-dire d'améliorer leurs performances à résoudre des tâches sans être explicitement programmés pour chacune. Il s'appuie, pour cela, sur des procédures, les algorithmes d'apprentissage, qui permettent à la machine de s'entraîner et d'apprendre à partir d'exemples qu'on lui présente.

Dans de nombreux cas d'utilisation de l'apprentissage automatique, les données sont naturellement décentralisées : les données des patients sont collectées et stockées par différents hôpitaux, les données sur notre environnement sont générées par des capteurs intégrés dans nos appareils personnels, etc. **Analyser l'ensemble de ces données peut permettre des avancées dans de nombreuses applications comme la médecine, ou encore le crowdsensing**, qui permet d'analyser et de cartographier des phénomènes d'intérêt sociétal (évolution de la qualité de l'air par exemple).

Or, il n'est pas toujours possible ou souhaitable de centraliser ces données au même endroit, pour des raisons de coût, de réglementation et/ou de confidentialité. C'est notamment le cas en santé où les données sont souvent cloisonnées dans différents « entrepôts » et où la réalisation d'études médicales multi-centriques (c'est-à-dire mettant en jeu les données de plusieurs centres de santé) est conditionnée à un processus de demande d'autorisation et une logistique longs et complexes.

¹ Pour "FEDerated MACHine Learning over the Internet".

² Voir Wikipedia (octobre 2022) : « L'apprentissage automatique (en anglais machine learning, littéralement 'apprentissage machine'), apprentissage artificiel ou apprentissage statistique est un champ d'étude de l'intelligence artificielle qui se fonde sur des approches mathématiques et statistiques pour donner aux ordinateurs la capacité d'« apprendre » à partir de données, c'est-à-dire d'améliorer leurs performances à résoudre des tâches sans être explicitement programmés pour chacune. Plus largement, il concerne la conception, l'analyse, l'optimisation, le développement et l'implémentation de telles méthodes. On parle d'apprentissage statistique car l'apprentissage consiste à créer un modèle dont l'erreur statistique moyenne est la plus faible possible. »

L'apprentissage fédéré, un changement de paradigme pour minimiser la collecte de données personnelles et protéger la confidentialité

L'apprentissage fédéré est récemment apparu comme un nouveau paradigme où **plusieurs entités avec des ensembles locaux de données entraînent des modèles d'apprentissage automatique de manière collaborative** tout en gardant leurs données décentralisées.

L'apprentissage fédéré permet de **minimiser la collecte de données personnelles**, et peut être combiné avec des techniques de **protection de la confidentialité** pour éviter que des informations sensibles ne soient révélées lors de l'apprentissage.

« FedMalin », un défi scientifique d'une ambition et d'une ampleur inégalées en matière d'apprentissage fédéré, en France et à l'international

Pionnier en France comme à l'international par son ampleur et son caractère largement pluridisciplinaire, **« FedMalin » est un Défi Inria co-porté par les chercheurs Aurélien Bellet et Giovanni Neglia** qui ambitionne de faire progresser la recherche sur l'apprentissage fédéré et de développer des cas d'utilisation concrets. Ceci, à travers un **consortium impliquant pas moins de 10 équipes-projet³** de l'institut et de ses partenaires, localisées dans 6 centres Inria⁴ et combinant des expertises en apprentissage automatique, systèmes distribués, confidentialité et sécurité, réseaux et médecine.

Les travaux de recherche porteront notamment sur **la confidentialité et l'équité, la consommation d'énergie, la personnalisation des modèles et les dépendances temporelles et spatiales**. FedMalin contribuera également au développement d'**outils open source** pour l'expérimentation de l'apprentissage fédéré et son déploiement dans le monde réel, et les utilisera pour des **applications concrètes en médecine et en crowdsensing**.

FedMalin se démarque sur la scène européenne et internationale par son large périmètre ainsi que par le savoir-faire et l'expertise de ses membres, qui va permettre de réaliser des avancées théoriques en apprentissage fédéré, mais aussi de rapidement les implémenter dans des bibliothèques logicielles open-source et de les appliquer à des problèmes réels.

Avant FedMalin, Mammals et Flamed

Le Défi FedMalin a partie liée avec un certain nombre de recherches menées au sein d'Inria. Parmi elles, deux « actions exploratoires », Mammals et Flamed, bénéficient également du soutien du Groupe La Poste. La première cherche à fournir des inférences à faible latence en exécutant, à proximité de l'utilisateur final, des modèles simples d'apprentissage automatique qui peuvent également exploiter une base des données locale. La seconde vise à appliquer l'apprentissage fédéré aux études médicales multi-centriques dans les hôpitaux.

Des applications concrètes nombreuses, au bénéfice d'un large public

Quoique mal connu encore, l'apprentissage fédéré constitue un **enjeu stratégique de premier plan** qui nécessitait une initiative d'ampleur comme FedMalin pour aider la recherche à franchir un cap décisif. Le sujet s'est développé fortement dans la communauté de l'apprentissage automatique ces deux dernières années mais commence seulement à se propager à d'autres domaines. De plus, les déploiements dans des applications réelles sont encore rares.

La recherche et les développements réalisés dans FedMalin dans le cadre d'une « science ouverte » pourront **bénéficier à un large public de citoyens et d'organisations publiques, ainsi qu'à l'ensemble**

³ Coati (CNRS, Inria, UCA) • Comete (CNRS, Inria, IPP) • Dyogene (CNRS, Inria, PSL) • Epione (Inria) • Magnet (CNRS, Inria, U. Lille) • Maracas (Inria, INSA Lyon) • Neo (Inria) • Spirals (CNRS, Inria, U. Lille) • Tribe (Inria) • Wide (Inria, U. Rennes).

⁴ Centre Inria d'Université Côte d'Azur, Centre Inria de l'Université de Lille, Centre Inria de l'Université de Rennes, Centre Inria de Paris, Centre Inria de Lyon, Centre Inria de Saclay.



de la communauté scientifique. FedMalin va aussi permettre de **consolider les projets existants d’Inria autour des applications médicales** et de renforcer les collaborations avec des acteurs de la santé, comme les hôpitaux du réseau Unicancer, le CHU de Lille et l’alliance G4. Du reste, cette coopération entre recherches de pointe en sciences et technologies du numérique et spécialistes des domaines applicatifs de l’apprentissage fédéré (médecins, radiologues...) est au cœur du modèle de l’apprentissage profond qui s’appuie sur des exemples « étiquetés » par des experts (de l’imagerie médicale, par exemple).

Les avancées de FedMalin apporteront au groupe La Poste, fort de plus de 450 experts en data et IA et acteur majeur de la santé, des **opportunités d’innovation dans les domaines de l’analyse des données de santé au bénéfice des patients et de la prévention à domicile** en faveur de l’accompagnement du vieillissement de la population. Des développements dans d’autres secteurs nécessitant une éthique de la donnée au service de la confiance numérique sont aussi envisagés.

Eu égard aux espoirs qu’il suscite en matière d’applications concrètes pour la santé, la mobilité, l’IoT..., le Défi FedMalin est emblématique d’une recherche « engagée » et « impliquée » au service de la société telle que le groupe La Poste, Inria et la Fondation Inria la promeuvent à travers leur partenariat.

A propos d’Inria

Inria est l’institut national de recherche en sciences et technologies du numérique. La recherche de rang mondial, l’innovation technologique et le risque entrepreneurial constituent son ADN. Au sein de 200 équipes-projets, pour la plupart communes avec les grandes universités de recherche, plus de 3 500 chercheurs et ingénieurs 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). Inria est labellisé Institut Carnot, confirmant sa volonté de rapprochement entre la recherche et le monde industriel.

A propos de la Fondation Inria

Fondation unique en France par son objet, la Fondation Inria veut contribuer à une transformation numérique à visage humain. Alors que le monde se digitalise, la fondation œuvre pour faire émerger et grandir des projets de recherche et d’innovation porteurs d’un impact positif pour les hommes, la société et la planète. Ceci, tout particulièrement dans quatre domaines clés où le numérique joue un rôle critique : la santé, l’environnement, l’éducation, la vie démocratique. À la fois think et do tank, la Fondation Inria propose un cadre simple et souple pour « donner du sens au numérique » en identifiant et déployant des programmes à impact qui permettent de mettre les sciences et technologies du numérique au service de l’intérêt général, sans angélisme ni solutionnisme technologique. Indépendante, la fondation agit grâce au soutien de ses partenaires et donateurs.

A propos du groupe La Poste

La Poste est une société anonyme à capitaux publics, filiale de la Caisse des Dépôts et de l’Etat. Le groupe La Poste est organisé en quatre branches : Services-Courrier-Colis, Grand Public et Numérique, GeoPost, La Banque Postale qui constitue avec sa filiale CNP Assurances le 11e bancassureur européen. La Poste distribue plus de 18 milliards d’objets par an dans le monde (lettres, imprimés publicitaires et colis), 6 jours par semaine. Attaché à sa présence territoriale, le groupe compte 17 000 points de contact (bureaux de poste, agences postales communales, relais poste commerçants) et 18 000 points d’accès à un service postal (Pickup, carrés pros, consignes ou encore drive colis). En 2021, le groupe La Poste a réalisé un chiffre d’affaires de 34,6 milliards d’euros, dont 41 % à l’international, et emploie près de 245 000 collaborateurs, dans 63 pays sur 5 continents dont 193 000 en France. Entreprise à mission depuis juin 2021, leader de la transition écologique et de la finance verte, La Poste ambitionne d’atteindre le « zéro émission nette » à horizon 2050. Avec son plan stratégique « La Poste 2030, engagée pour vous », l’entreprise publique se fixe l’ambition de devenir la première plateforme européenne du lien et des échanges, humaine et digitale, verte et citoyenne, au service de ses clients dans leurs projets et de la société tout entière dans ses transformations.

Contacts presse

- > Inria : Laurence Goussu - laurence.goussu@inria.fr - 01 39 63 57 29
- > Fondation Inria : Flavie Dubois-Mazeyrie - flavie.dubois-mazeyrie@fondation-inria.fr - 01 80 49 45 46
- > La Poste : Patrice Lemonnier - patrice.lemonnier@laposte.fr - 01 55 44 25 35