

4 novembre 2025

## Communiqué de presse

### *Et si votre voix aidait à préserver les langues de France ?*

***Inria et ses partenaires engagent une démarche collaborative pour préserver la diversité linguistique et l'intégrer dans l'intelligence artificielle et les prochaines technologies vocales.***

*En dehors du français, les langues de France sont très peu présentes dans les technologies vocales et numériques. Pour changer la donne, Inria lance une campagne nationale et participative de collecte de voix. Cette initiative inédite permettra de préserver la diversité linguistique et de donner aux langues de France leur place dans l'intelligence artificielle.*

### **Un défi majeur : ne laisser aucune langue de côté**

Le projet COLaF (Corpus et Outils pour les Langues de France), porté par Inria, veut relever un défi ambitieux :

- Constituer des **corpus textuels et oraux** dans toutes les langues parlées en France, y compris la langue des signes française.
- Développer des **outils numériques libres** pour la recherche, l'éducation, l'industrie et la société civile.
- **Valoriser la diversité linguistique dans toute sa richesse** : français hexagonal et ultramarin, langues régionales (breton, occitan, alsacien...), créoles, langues autochtones et d'immigration, français d'apprenants, etc.

⇒ **Objectif : outiller les langues de France pour le XXI<sup>e</sup> siècle.**

Ce défi est réalisé dans le cadre d'une collaboration entre des laboratoires universitaires, des institutions culturelles, des projets numériques ouverts et bénéficie du soutien du Programme National de Recherche en Intelligence Artificielle (PNRIA).

### **Une collecte participative via Common Voice**

Pour atteindre cette ambition, COLaF s'appuie sur plusieurs initiatives ouvertes de collecte vocale. Le principe est simple :

- Toute personne peut contribuer en quelques minutes en enregistrant sa voix depuis son téléphone ou son ordinateur.
- Les données collectées, placées dans le domaine public, constitueront une ressource inestimable et librement réutilisable pour alimenter la recherche et l'innovation en reconnaissance et synthèse vocale.

## Zoom sur “Parole Spontanée” : une première mondiale en alsacien

L’alsacien est la première langue de France testée dans le dispositif inédit : “**Parole Spontanée**”. Contrairement à l’approche classique de Common Voice qui se base sur la lecture de phrases pré-écrites, chaque participant pourra s’exprimer librement sur un sujet donné. Cette approche ludique et efficace vise à accélérer la collecte pour de nouvelles langues.

Derrière ce projet pilote, de nombreux acteurs partenaires :

- L’équipe-projet Multispeech, commune au CNRS, à Inria et à l’Université de Lorraine, au sein du Centre Inria de l’Université de Lorraine et du Loria (Laboratoire Lorrain de recherche en informatique et ses applications).
- L’Université de Strasbourg (Faculté des langues et laboratoire LiLPa - Linguistique, Langues et Parole, en charge de la traduction intégrale de l’interface en alsacien)
- Le Voice Lab
- La Délégation générale à la langue française et aux langues de France, service rattaché au Ministère de la Culture.
- Le Programme National de Recherche en IA

D’autres langues suivront : breton, picard, créole,...

## Le Voice Lab : un écosystème engagé pour une IA francophone souveraine

Partenaire clé de cette campagne, Le Voice Lab mobilise chercheurs, entreprises et institutions pour développer des solutions vocales **souveraines, éthiques et performantes**. Les données collectées seront mutualisées via son **Speech Data Hub**, une plateforme ouverte dédiée à la valorisation des ressources vocales.

## Un projet citoyen et culturel

Contribuer à COLaF, c’est :

- **Préserver et transmettre** le patrimoine linguistique français aux générations futures.
- **Renforcer** leur visibilité dans le monde numérique.
- **Contribuer au développement d’une intelligence artificielle plus juste**, plus inclusive, capable de s’exprimer dans la langue de chacun
- Proposer des outils adaptés à chaque langue, pour répondre à certains besoins du quotidien.

Rejoignez le mouvement :  
enregistrez quelques minutes de votre voix et faites entrer votre langue dans l’IA !

Pour participer : <https://colaf.huma-num.fr/common-voice/>

## Liens utiles

En savoir plus sur le projet COLaF : [www.colaf.huma-num.fr/#team](http://www.colaf.huma-num.fr/#team)

Participer au projet Parole Spontanée : <https://colaf.huma-num.fr/common-voice/>

## Contacts presse

Inria : Véronique Poirel - [veronique.poirel@inria.fr](mailto:veronique.poirel@inria.fr) | Magalie Quet - [magalie.quet@inria.fr](mailto:magalie.quet@inria.fr)



**A propos d’Inria :** Inria est l’institut national de recherche en sciences et technologies du numérique et a la responsabilité depuis janvier 2024 de l’Agence de programmes Numérique pour renforcer les dynamiques collectives de l’enseignement supérieur et de la recherche. La recherche de rang mondial, l’innovation technologique et le risque entrepreneurial constituent son ADN. Au sein de 220 équipes projets, pour la plupart communes avec les grandes universités de recherche, plus de 3 800 scientifiques y explorent des voies nouvelles, souvent dans l’interdisciplinarité et en collaboration avec des partenaires industriels pour répondre à des défis ambitieux. Institut technologique, Inria soutient la diversité des voies de l’innovation : de l’édition open source de logiciels à la création de startups technologiques (Deeptech).