Alejandrina Cristia

Paris, France • alecristia@gmail.com • LinkedIn profile

Senior Research Scientist at the intersection of AI/ML and sensitive data

Over more than a decade at CNRS, I've taken work from niche idea to widely used tool: field-testing recording hardware across four continents, supervising the development of state-of-the-art open-source speech AI, building the datasets and annotation pipelines to train and benchmark models, and working through the ethics and legal aspects of sensitive child-centered recordings. I've also done the operational work: hiring 20+ people, writing onboarding and training documentation, managing budgets across funding lines, reviewing code, and communicating with stakeholders. I coauthored 150+ peer-reviewed papers across disciplines, leading to international recognition in speech technology (2025 ISCA fellow) and development economics (J-PAL Invited Researcher). Now I'm seeking to move from research tools to serving millions with AI and data products that are safe, fair, and useful for very diverse users, including children and under-served language communities.

Selected Positions and Education

2013–2026 | Researcher → Senior Researcher, *CNRS & École Normale Supérieure, Paris, France*
2011–2012 | Postdoctoral Fellow, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands
2009 PhD in Linguistics, Purdue University, USA
2004 Profesorado de Letras, Universidad Nacional de Rosario, Argentina

Selected Achievements

AI/ML and speech technology

- Co-organized the DIHARD I & II speaker diarization challenges and the JSALT 2019 diarization track.
- Voice Type Classifier (VTC): state of the art on child-centered audio since 2020, including against commercial alternatives. Co-supervised BabyHuBERT, a multilingual SSL model pretrained on 13k hours of child-centered speech, with SOTA on voice type classification (Charlot et al., 2026).

AI policy, privacy, and data protection

- Co-authored the *Data Protection Handbook for Long-Form Recording Research*, reviewing legislation in 28 countries and translating it into practical guidance for collecting, storing, and reusing sensitive child-centered recordings.
- Pioneered data-protection practices for child-centered audio research at CNRS with Data Protection Officers; developed data-management plans, DPIAs, legal and ethical frameworks, and workflows that embed them across the data lifecycle.
- Led governance of BabyHuBERT, including a licensing approach to mitigate risks; within the ERC ethics review of novel AI methods, established an ad hoc ethics board; advised ethics boards and international archives on reuse of sensitive data (for example, access rules set by aboriginal leadership).

Hardware product development

- Led an international initiative to enable safe, comparable audio collection from infants across diverse cultural, climatic, and economic settings: evaluated commercial and custom recorders against performance, usability, child safety, family acceptability, and cost, and ran iterative field pilots in Argentina, Namibia, Bolivia, Vanuatu, and Paris.
- Identified key constraints (short hardware lifecycles, costly custom devices, no one-size attachment design) and defined a scalable strategy based on selection principles rather than a fixed product; produced various training materials, including youtube videos.

Datasets and open-source software

- Developed annotation taxonomies and innovated crowd-sourced annotation for sensitive child-centered recordings.

- Supervised the creation of multi-level acoustic and text datasets for developing and benchmarking AI models: stratified sampling that audits and minimizes demographic and acoustic bias, train/dev/test protocols with grouping variables to prevent contamination, and quality control from inter-annotator reliability to derived metrics.
- Coordinated development of open-source research software (version control, issue tracking), with documentation, educational materials, and user support for teams inside and outside my own.

Management and operations

- In addition to other grants, secured and led a competitive €2 million, five-year European Research Council Consolidator Grant for AI development on long-form audio recordings from children and their families collected in multiple countries.
- Led an interdisciplinary team of up to 12 staff plus external collaborators, coordinating scientific discovery, AI development, participant protection, and risk management; managed milestones, deliverables, and reporting across an international network; built and led consortia of researchers and practitioners across multiple continents.
- Built the team: scoped needs and divided expertise into in-house, outsourced, and hired; designed a hiring system and recruited 20+ technical, administrative, and research staff; built onboarding and offboarding (contracts from one week to several years) and role-specific training, including for annotators and ML engineers; coached junior researchers; ran performance management, including cases ending in termination; managed budgets across funding lines and trained administrative staff.
- Communicated complex technical, ethical, and legal issues to non-specialists: produced policy and governance briefings and reports for institutional leadership, ethics reviewers, practitioners, international organizations, and community representatives; co-created and taught courses on the societal implications of research and AI (*Cognitive Science in the Real World, AI for Social Good*).

Other expertise

- Evidence synthesis and statistical modeling: systematic reviews and meta-analyses with bias auditing; a hybrid meta- and mega-analytic method to impute missing data; hierarchical models, simulations, DAGs, and Bayesian modeling; results translated for non-statistical partners.
- Speech perception, production, and cognitive development: pioneering measurement of individual variation in infant speech perception; acoustic-phonetics of child and child-surrounding speech; cognitive measures that are fair across cultures, languages, and multilingual settings.

Additional Information

- International: lived and worked in Argentina, the UK, the USA, the Netherlands, and France; fieldwork in Bolivia and Papua New Guinea; advised teams in Timor-Leste, the Solomon Islands, India, and Ghana; visiting fellow at the Max Planck Institute for Comparative Cultural Psychology (2022) and Aarhus University's Interacting Minds Center (2025, 2026).
- Effective giving: Giving What We Can 10% pledge since 2020; co-founded Don Efficace (2023–2025) and Générosité à Impact (2025).
- Languages and citizenship: native Spanish; native-like English and French; conversational Italian. Argentinian, Italian, and French citizenships.