Dr. Arjhun Swaminathan

Technical AI Governance & Safety researcher building privacy-preserving ML systems for high-stakes data; CIPP/E + AIGP; translating AI risk, privacy, and regulatory constraints into technical evaluations and system design.

German citizen, EU work authorization **Languages:** German (B1) · English (C2) · Tamil (C2)
GitHub: Arjhun-alt | ORCID: 0009-0004-3223-4538

Selected Publications

PP-GWAS — *Nature Communications*
Built an end-to-end pipeline for privacy-preserving distributed genetic association studies; ~2× faster than prior baselines with lower resource usage.

TEA — *SaTML 2026*
Built a targeted hard-label black-box attack that uses edge information from the target image; reaches a 60% distance reduction in 251 queries on average across four models, against 845 for the next fastest method.

Trustworthy MCP — *ICML 2026 · Position*
Analyzed MCP as a protocol-level intervention point for responsible agentic AI, connecting system architecture with trustworthy-AI and governance requirements.

Experience

since 10.2026 — Remote
Independent Researcher
Technical AI Safety | funded by Coefficient Giving
Full-time research and training in technical AI safety.

from 01.2027 — Remote
AI Safety Camp 12
Project Lead | Technical AI Governance
- Selected to lead a team modeling compute-permit prices and audit rules.

since 09.2026 — Remote
Supervised Program for Alignment Research (SPAR)
Research Fellow | Technical AI Governance
- Study how agent interoperability standards affect intentional loss-of-control risk.

04.2026–09.2026 — Remote
Safe AI Germany (SAIGE)
Research Fellow | Technical AI Governance
- Extended a formal model of risk-weighted compute permits to audits facing hidden compute and manipulated evaluations.
- Led the resulting paper (under review) and co-authored a public policy brief.

10.2022–09.2026 — Tübingen, Germany
University of Tübingen
Scientific Researcher | Privacy-Preserving Machine Learning
- Published eight peer-reviewed papers.
- Built and open-sourced end-to-end implementations of privacy-preserving ML methods, adopted by external partners.
- Supervised a master's thesis on quantum federated learning with fully homomorphic encryption through to publication at CYBERSEC 2026.
- Taught and mentored 75+ graduate students across a lecture with tutorials, a seminar, and a practical course.

Education

10.2022–07.2026 — University of Tübingen
Doctor of Philosophy (PhD) | *Computer Science* — Grade – 1.0 (100%)
Thesis – *Privacy by Design: From Distributed Learning to Post-Deployment Risks*

11.2020–09.2022 — University of Tübingen
Master of Science (MSc) | *Mathematical Physics* — Grade – 1.53 (91%)
Thesis – *Supersingular Isogeny Diffie-Hellman Key Exchange*

Grants & Awards

2026
Career Transition Grant, Coefficient Giving
To fund full-time independent research and training in technical AI safety.

Certifications

Artificial Intelligence Governance Professional (AIGP) | *EU AI Act* | *ISO/IEC 42001* | *Risk Management*
Certified Information Privacy Professional/Europe (CIPP/E) | *GDPR* | *ePrivacy Directive*

Skills

Programming — Python, Git, Linux
ML/Data — PyTorch, torchvision, scikit-learn, NumPy, SciPy, pandas