

Research Statement

Xiaoming Fu

Professor (W3) of Computer Science
Head, Computer Networks Group
Institute of Computer Science
University of Göttingen

Research Vision and Overview

My research focuses on the design and orchestration of AI-enabled socio-technical infrastructures for sensing, knowledge discovery, and decision support in complex societal systems. As artificial intelligence, pervasive sensing, edge-cloud computing, and large-scale digital connectivity increasingly reshape both technological systems and human society, I investigate how digital infrastructures can support intelligent, sustainable, and resilient societies.

A central conviction underlying my work is that digital infrastructures are no longer merely communication substrates. They have evolved into intelligent systems that sense the physical and social world, transform heterogeneous data into actionable knowledge, and support autonomous decision-making, collective coordination, and evidence-based policy. This transformation raises fundamental research questions: *How can distributed infrastructures efficiently support AI workloads at scale? How can AI systems become resource-aware, trustworthy, and resilient? And how can large-scale sensing and data infrastructures generate actionable knowledge for societal transformation?*

Over 24 years at the University of Göttingen, I have built an internationally recognized research program at the intersection of AI, networking, distributed systems, data-centric Internet architecture, and computational social science. My group has graduated more than forty PhD students and secured competitive funding from the European Commission (FP6, FP7, H2020, Horizon Europe), German federal/state programs, and multiple international collaborations.

Core Research Themes

AI Infrastructure and Distributed Intelligence

A major pillar of my research concerns the co-design of AI systems and digital infrastructures, enabling scalable and efficient intelligence across heterogeneous edge-cloud environments.

Recent work focuses on efficient large-model inference under stringent resource constraints. We developed memory-efficient KV cache optimization methods (IEEE INFOCOM 2026) and small-model compensation techniques for KV cache compression (NeurIPS 2025), enabling large language model inference on memory-constrained edge hardware while preserving model quality. These methods substantially reduce memory footprint, latency, and energy consumption, making advanced AI services increasingly deployable at the edge.

Beyond inference, my group studies distributed AI training and cooperative intelligence at scale. The Barre system (USENIX ATC 2025) introduces programmable congestion control tailored to GPU cluster communication patterns in distributed AI training. ReSCOM (AAMAS 2025 Best Paper Runner-Up; TMC 2026) develops reward-shaped curriculum learning for efficient communication in decentralized multi-agent systems, laying foundations for agentic AI systems.

These efforts advance the infrastructure required for the next generation of large-scale distributed intelligence.

Edge-Cloud Systems and Autonomous Resource Orchestration

My earlier work established important foundations in mobile edge computing, computation offloading, and network softwarization.

Our IEEE/ACM ToN 2016 paper on multi-user computation offloading and the D2D Fogging framework (IEEE JSAC 2016) introduced energy-aware and incentive-compatible task offloading mechanisms before mobile edge computing became mainstream. These works helped shape later developments in edge intelligence and distributed computing.

In network softwarization, we developed NFVnice (SIGCOMM 2017) and resilient service-chain orchestration methods (CoNEXT 2018), contributing to scalable resource management for virtualized network services.

These foundations now extend into the Horizon Europe CODECO project (2023–2026), which investigates autonomous orchestration across the edge-cloud continuum using cognitive control and decentralized intelligence.

Data-Centric Networking and Internet Architecture

A long-standing research direction concerns data-centric Internet architectures, particularly Information-Centric Networking (ICN).

Between 2013 and 2019, I coordinated major EU projects including GreenICN, ICN2020, MobileCloud, and CleanSky. These projects advanced data-centric networking and cloud infrastructures from foundational architectural research toward practical deployment, with improved resource utilization and energy efficiency for large-scale distributed and mission-critical systems such as disaster management.

Our work on ICN-enabled SDN management (ACM ICN 2014 Best Paper Award) and content-oriented publish/subscribe systems (ACM/IEEE ANCS 2011) contributed significantly to data-centric communication paradigms and informed Internet standardization, including IETF RFC 8884.

This research has gained renewed relevance in the AI era, where named and reusable data objects increasingly serve as the foundation for scalable AI inference, multimodal knowledge retrieval, and digital-twin-driven knowledge discovery and decision support.

Secure and Resilient Digital Systems

Security, privacy, and resilience are essential properties of trustworthy digital infrastructures.

BotFinder (CoNEXT 2012) pioneered machine-learning-based bot detection without deep packet inspection. ADA (INFOCOM 2020) extended anomaly detection to infrastructure-scale log analysis. Our work on mobility privacy (WWW 2017) quantified fundamental limitations of anonymization in aggregate mobility traces.

My group has also long studied resilient communications for emergency and disaster scenarios. CNS: Content-oriented Notification Service for Managing Disasters (ICN 2016) introduced a role-aware communication architecture enabling efficient coordination among first responders under disrupted network conditions. This work established an early socio-technical perspective on resilient digital infrastructures by integrating networking mechanisms with organizational coordination during crisis response.

Building on this foundation, the Horizon Europe COVER project (2023–2026) investigates secure and resilient UAV-assisted emergency communication systems and AI-supported disaster response.

Computational Social Science and AI for Society

A long-standing interdisciplinary dimension of my research applies network science, computational social science, and AI to understand human behavior, social systems, and societal transformation.

A central methodological principle in this work is that meaningful insights from large-scale behavioral data emerge not from data-driven prediction alone, but from the interaction among computational analytics, sociological theory, and predictive modeling. This perspective was articulated in our conceptual work on theory-guided big data research (Journal of Chinese Sociology, 2019), which argues for a triadic dialogue among big data analysis, theory construction, and predictive inference.

Building on this perspective, our work on social network evolution developed predictive models linking structural network dynamics to relationship formation and interaction patterns. Studies on triadic closure and tie-strength evolution in social networks (IEEE TKDE 2015; ACM TKDD 2018) provided theoretical insights into social link formation, while cross-platform consumer behavior analysis from large-scale mobile shopping data (WWW 2018) demonstrated how digital traces reveal behavioral dynamics at scale.

Subsequent work extended toward socioeconomic inference and human mobility analytics. Research on socioeconomic status prediction from mobile phone and social media data (International Journal of Data Science and Analytics 2020; Expert Systems with Applications 2022) and mobility prediction across urban environments (DASFAA 2023; IEEE Transactions on Computational Social Systems 2024) showed how large-scale digital behavioral data can uncover latent socioeconomic structures and mobility patterns.

These studies established methodological foundations for my more recent work on AI for sustainability, social equity, and policy intelligence.

AI for Sustainability and Human-Centered Transformation

An increasingly central dimension of my research applies AI, sensing, and large-scale data analytics to sustainability transitions and societal resilience.

In a German Excellence Initiative, I led research on land-use data science, integrating remote sensing, agricultural

sensing, and soil-quality analytics for sustainable land management. Related work on rural socioeconomic development in Thailand and Vietnam using satellite imagery (Computer Communications 2023) demonstrated how AI-enabled sensing infrastructures can support sustainable development assessment and policy planning in Global South contexts.

My current initiatives address AI-assisted smart energy transformation, resilient urban systems, and climate adaptation. In collaboration with humanitarian stakeholders including the Red Cross, Technisches Hilfswerk (THW) and municipalities, I contribute to AI-assisted flood management and disaster response, supporting data-driven resilience planning.

I also study how AI can improve social equity and human wellbeing. Research on online shadow education (Education and Information Technologies 2024) identified mechanisms through which digital learning platforms can reduce or exacerbate educational inequality. More recently, the Intelligent Innovation Dataset on Scientific Research Outcomes and Patents (Journal of Social Computing 2025) provides a large-scale data infrastructure for analyzing scientific outputs, innovation ecosystems, and translational research pathways, enabling AI-driven knowledge discovery for research evaluation and innovation policy.

The ENLIGHT project DISRUPT (2024–2027) further extends this agenda toward digital innovation for urban and periurban transitions. More broadly, I am expanding this line of research toward computational humanities and historical AI. Using large-scale digitized historical records, we reconstruct elite social networks and mobility trajectories across major pre-modern empires to enable data-driven analysis of institutional openness, social mobility, and governance resilience across civilizations.

These activities reflect my broader vision: AI-enabled infrastructures should not only optimize technical systems but also generate actionable knowledge that supports sustainable governance, equitable development, and evidence-based policy.

Research Impact and Community Leadership

My work has appeared in leading venues including ACM SIGCOMM, ACM CoNEXT, IEEE INFOCOM, USENIX NSDI, USENIX ATC, AAAI, AAMAS, NeurIPS, and flagship journals including IEEE/ACM Transactions on Networking, IEEE Transactions on Mobile Computing, IEEE JSAC, IEEE TCSS, IEEE TKDE, and ACM TKDD.

Recognition of this work includes election as IEEE Fellow (2022), membership in Academia Europaea (2018), and election to acatech, the German National Academy of Science and Engineering (2024).

I actively contribute to scientific leadership and community building across networking, AI, and computational social science. I currently serve as Editor-in-Chief of ACM Transactions on Social Computing (TSC), helping shape the evolving field of AI-driven social computing and computational social science.

In addition, I have served as TPC Co-Chair or General (Vice) Co-Chair for major international conferences including IEEE ICNP, IEEE ICDCS, ACM ICN, ACM MobiCom, and as a TPC member of ACM MobiCom, CoNEXT, IEEE INFOCOM, AAAI, and other leading conferences. I also serve or have served on the editorial boards of multiple IEEE/ACM journals and magazines. Through these roles, I contribute to shaping research directions across networking, distributed systems, AI infrastructure, and socio-technical computing.

Future Research Directions

Looking ahead, three directions define the frontier of my research programme:

- *Sustainable AI inference infrastructure*: As foundation models move from centralized cloud environments toward edge deployment, resource efficiency, latency, and energy consumption become fundamental bottlenecks. I will develop sustainable AI-serving architectures for heterogeneous edge-cloud systems.
- *Agentic socio-technical systems*: Future AI systems will increasingly consist of autonomous agents interacting over networks and embedded within human organizations. I will study communication, trust, coordination, and governance mechanisms that enable robust multi-agent systems in complex real-world environments.
- *AI for Knowledge Discovery and Policy Intelligence*: I aim to develop AI-enabled infrastructures that transform heterogeneous large-scale data into actionable knowledge supporting intelligent decision-making and evidence-based policy across energy systems, climate resilience, smart cities, scientific innovation, education, and even historical institutional analysis.

These directions extend my long-term vision of building intelligent digital infrastructures that bridge foundational systems research with societal transformation.