

Challenges of Designing Inclusive and Value-Aligned AI in the Public Sector

Nitin Sawhney

Aalto University & University of the
Arts Research Institute, Helsinki
nitin.sawhney@uniarts.fi

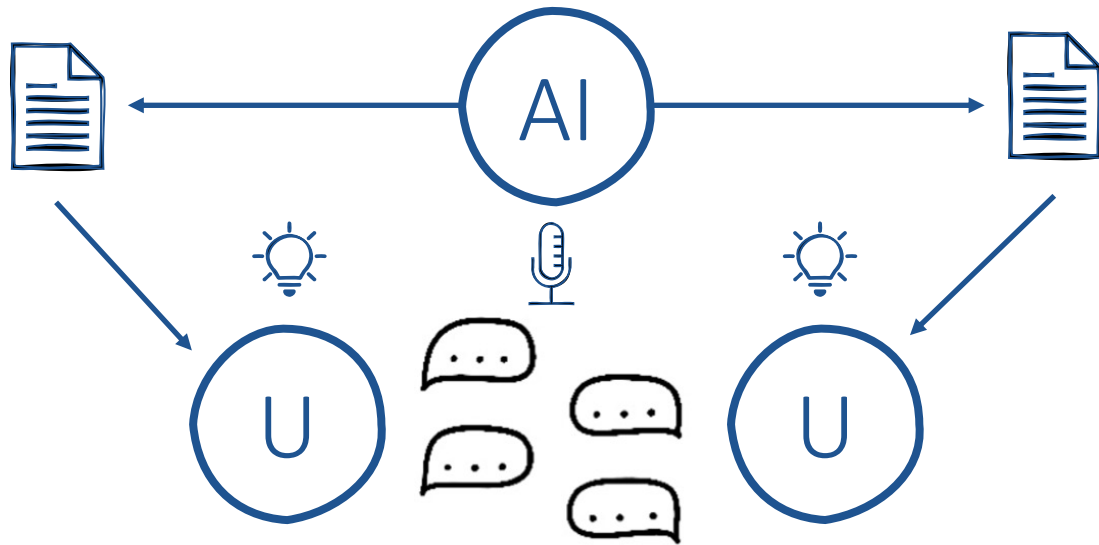
- Examining Discourses for Shared and Contested Public Values
- Human-AI Collaboration for Trustworthy Public AI Services
- Role of AI Regulatory Sandboxes

Examining discourses of shared and contested values embedded in the EU AI Act and Public AI Services



- Gonzalez Torres, A. P., Kajava, K., & Sawhney, N. 2023. Emerging AI Discourses and Policies in the EU: Implications for Evolving AI Governance. *Communications in Computer and Information Science*, Springer.
- Kajava, K. and Sawhney, N. 2023. Language of Algorithms: Agency, Metaphors, and Deliberations in AI Discourses. Lindgren, S. *Handbook of Critical Studies of Artificial Intelligence*. Edward Elgar Publishers.

Enhancing Conversations in Migrant Counseling Services: Designing for Trustworthy Human-AI Collaboration



Examining existing values and practices of city counselors from conversations with migrant users. Role of distributed knowledge (& values) emerging in human-AI interaction?

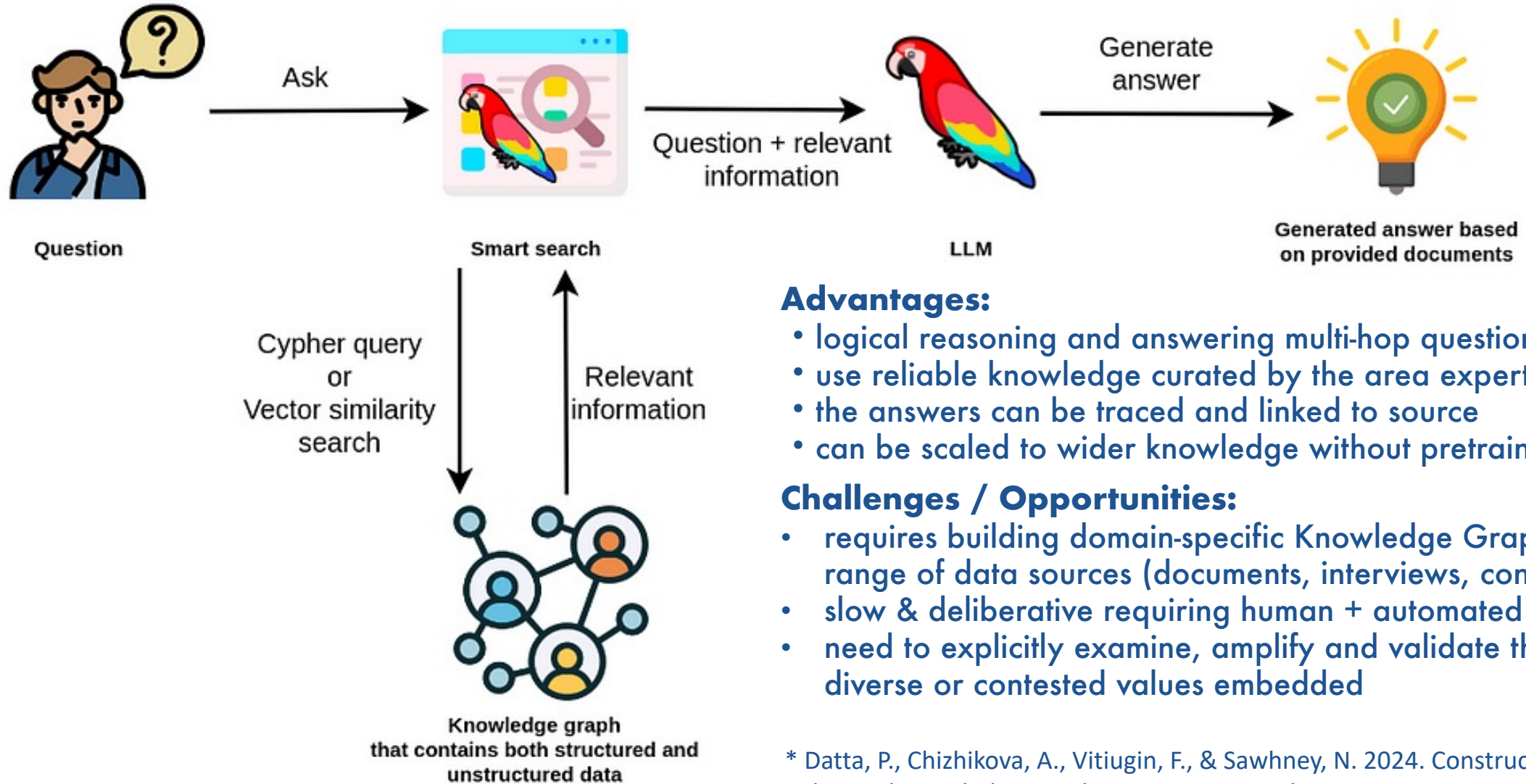


Design research & prototyping at Aalto University, Lucy Truong's thesis research with the City of Espoo (2023)



* Truong, L., Lee, S., & Sawhney, N. 2024. Enhancing Conversations for Migrant Counseling: Designing for Trustworthy AI-mediated Collaboration between Migrants and Service Advisors. *ACM Conference on Computer-Supported Cooperative Work and Social Computing*, San José, Costa Rica, November 9-13, 2024.

Embedding Public Values in Knowledge Graphs?



Advantages:

- logical reasoning and answering multi-hop questions
- use reliable knowledge curated by the area experts
- the answers can be traced and linked to source
- can be scaled to wider knowledge without pretraining

Challenges / Opportunities:

- requires building domain-specific Knowledge Graph from a range of data sources (documents, interviews, conversations)
- slow & deliberative requiring human + automated process
- need to explicitly examine, amplify and validate the nature of diverse or contested values embedded

* Datta, P., Chizhikova, A., Vitiugin, F., & Sawhney, N. 2024. Construction of Hyper-Relational Knowledge Graph Using Pre-Trained Large Language Models. *AAAI Conference on Artificial Intelligence, Workshop on Graphs and more Complex structures for Learning and Reasoning*, February 20-27, 2024, Vancouver, Canada.

AI Regulatory Sandboxes for Public Sector Innovation

*A concrete and controlled **framework** set up by a competent authority which offers providers or prospective providers of AI systems the **possibility to develop, train, validate and test**, where appropriate in real world conditions, **an innovative AI system**, pursuant to a sandbox plan for a **limited time under regulatory supervision**.*

EU AI Act: Article 3(55)

** However, EC Proposal article 52(4) indicates that participants in AI regulatory sandboxes would remain **liable** for any harm inflicted on third parties as a result of experimentation in the sandbox environment.*

ENGAGING PUBLIC VALUES IN AI REGULATORY SANDBOXES

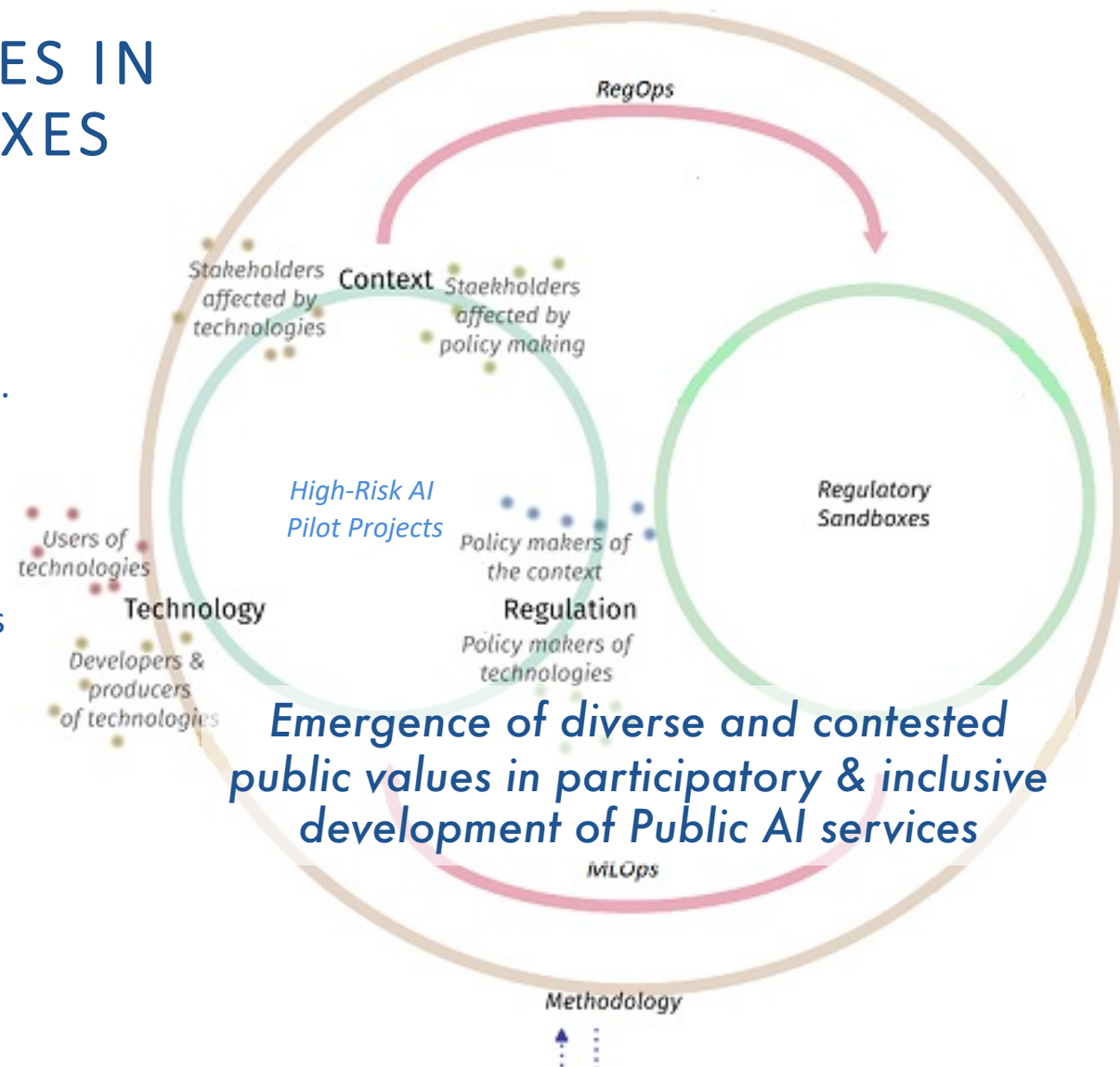
Conduct pilots of AI Public Sector services using AI Regulatory Sandboxes, integrated platforms and participatory infrastructures, in focused domains like public health, municipal services etc.

Engage with multiple stakeholders:

- Regulators, public administration, providers, developers, domain experts, SMEs and citizens and civil society groups.

Expected outcomes:

- Developing platforms, guidelines, policies and best practices to support sandboxes
- Provide the space for interactions and mutual collaboration though out the AI lifecycle
- Examine the risks, limitations and possibilities of legislation and technological innovation



* Gonzalez Torres, A. P. and Sawhney, N. 2023. Role of AI Regulatory Sandboxes and MLOps for Finnish Public Sector Services. *Special issue of The Review of Socionetwork Strategies (RSS)*, Springer.