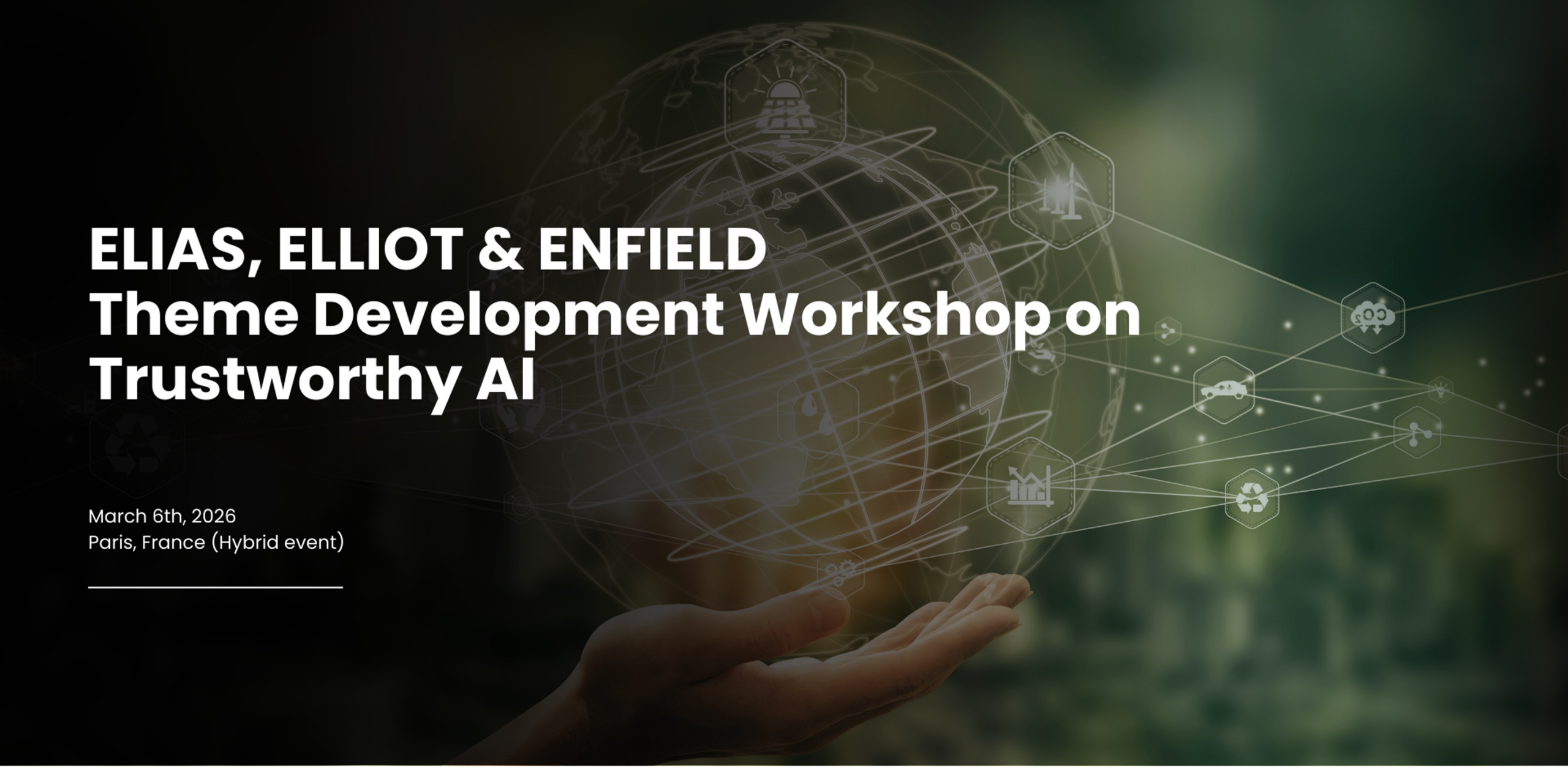


ELIAS, ELLIOT & ENFIELD Theme Development Workshop on Trustworthy AI

March 6th, 2026
Paris, France (Hybrid event)

Funded by the European Union

ELIAS

elliot
MULTIMODAL AI MADE IN EUROPE

ENFIELD

TDW on Trustworthy AI

Theme Development Workshops

March 6th, 2026
Paris, France (Hybrid event)

Human Perception of AI Trustworthiness

Dr. Sebastian Heil
Senior Researcher@TUC
ENFIELD TAI Task Co-Leader
sebastianheil

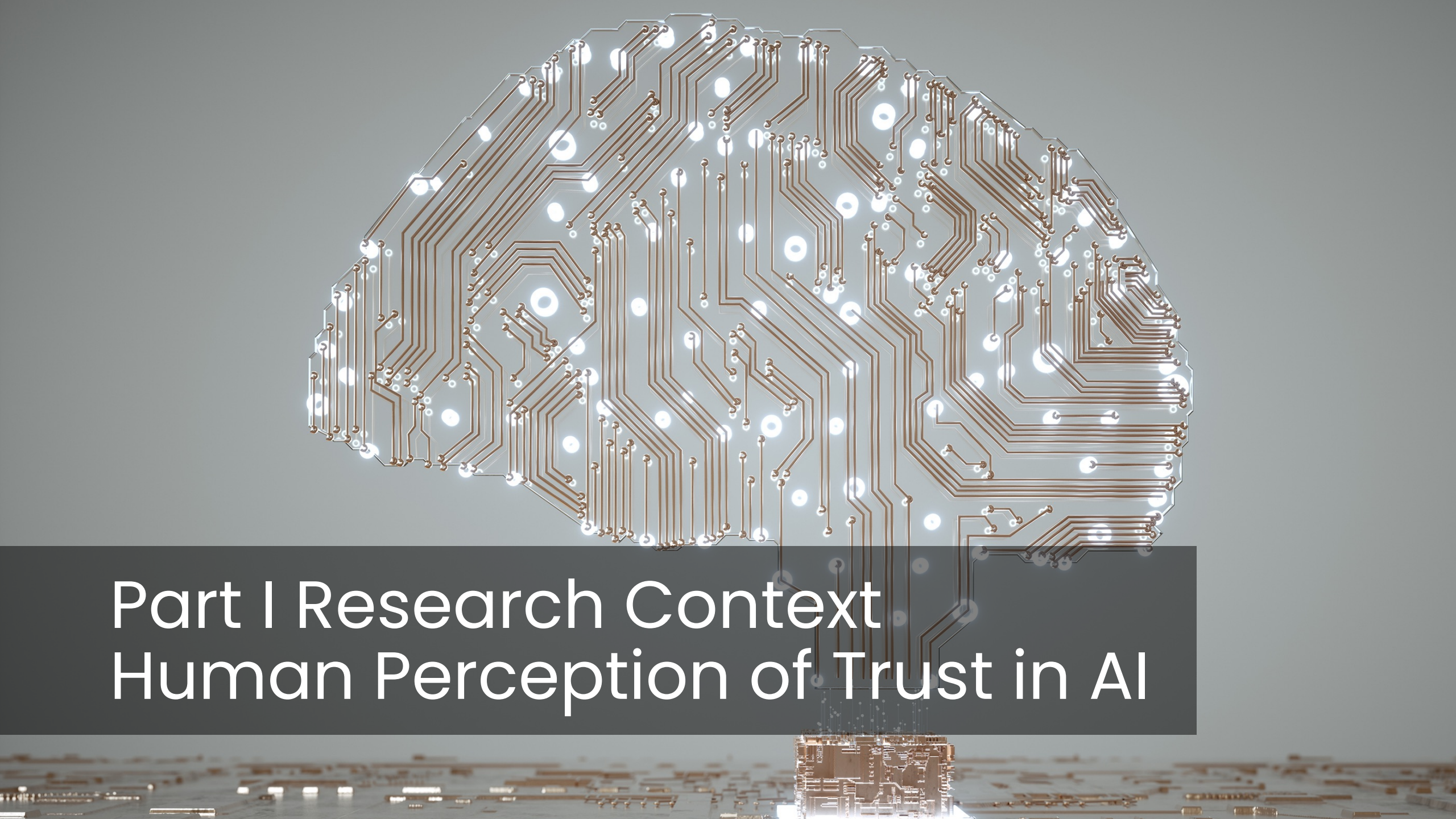
Funded by the European Union

Outline

RESEARCH CONTEXT

THEORETICAL FOUNDATIONS

CURRENT RESEARCH



Part I Research Context
Human Perception of Trust in AI

ENFIELD T2.4 RT3 User Perception and Expectation

Scientific Challenge

Understanding how users perceive AI from the perspective of trust and confidence in the technology, which in turn includes multiple possible perspectives (including accuracy, reliability, safety, security).

Expected result

Factors that motivate or impact user trust in AI technologies and their relative weightings in different contexts.

Involved ENFIELD Partners

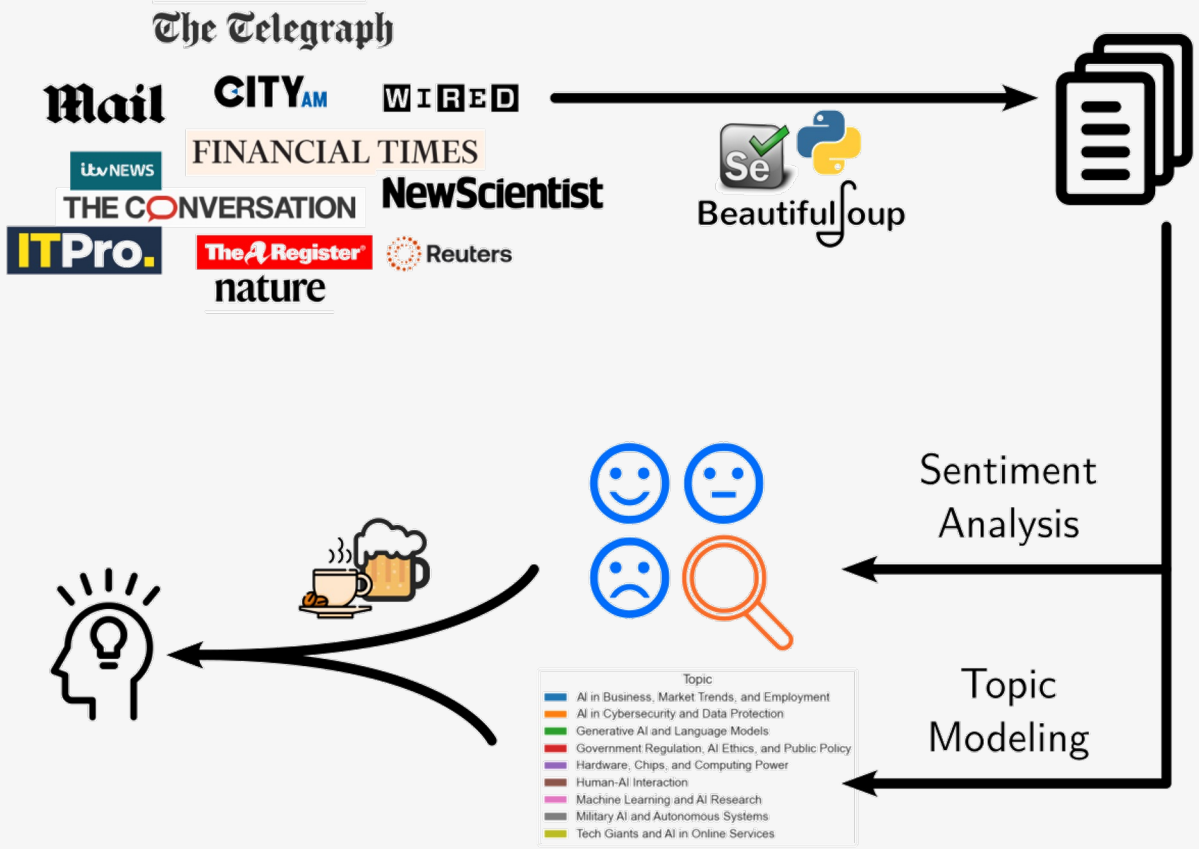
Public Perceptions of Trustworthy AI: Insights from a Longitudinal Study of UK News Media

Large-scale, longitudinal mapping of UK AI news discourse

- 7,691 articles from 2013 to 2024, from 12 UK news outlets
- Mainstream, business, scientific and technology themes
- Diverse range of viewpoints and readerships, aiming to capture a broad spectrum
- Descriptive statistics, sentiment analysis, and topic modeling

Findings

- Generally optimistic but increasingly critical: reflects a maturing public discourse
- Shifting from celebrating technical achievements to expectations in accountability



Murphy, T., Furnell, S., Heil , S. and Carpent, X. (2025). Public Perceptions of Trustworthy AI: Insights from a Longitudinal Study of UK News Media. 19th IFIP International Symposium on Human Aspects of Information Security & Assurance.

Part II
Theoretical Foundations

About Trust and Trustworthiness

Trust is earned in drops and lost in buckets.

Kevin Kelly, founding executive editor of Wired magazine

Trust is viewed as:

(1) a set of specific **beliefs** dealing with benevolence, competence, integrity, and predictability (trusting beliefs);

(2) the willingness of one party to **depend** on another in a **risky situation** (trusting intention); or

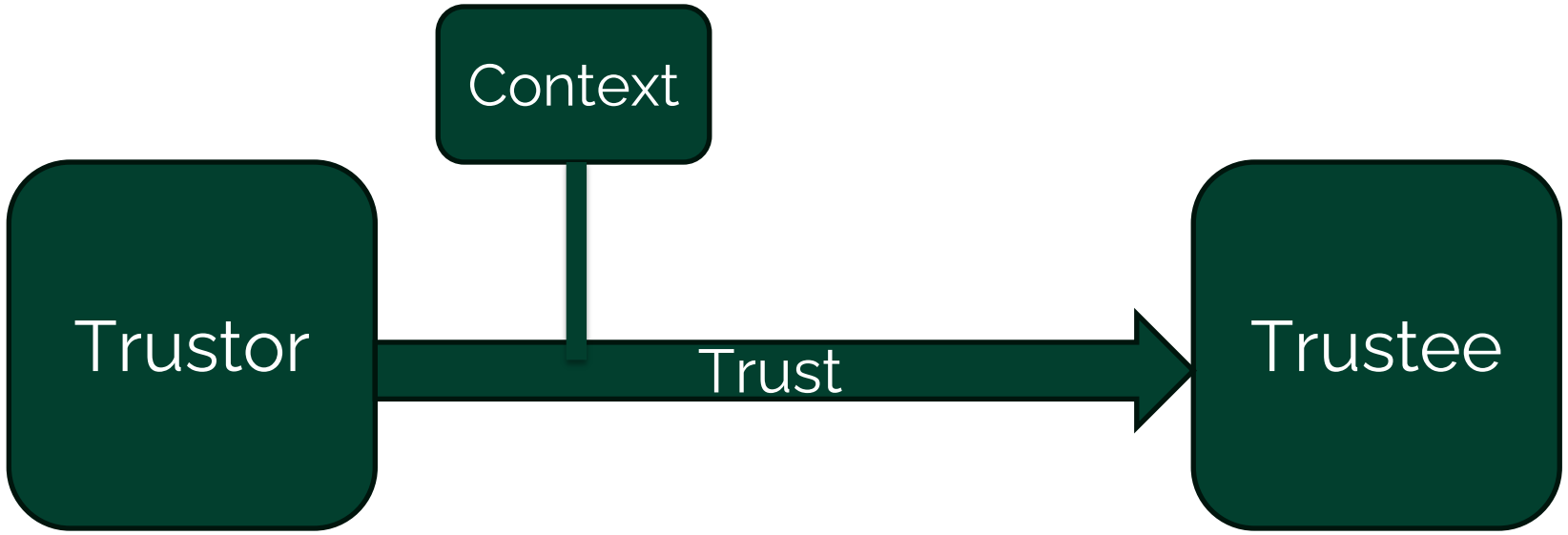
(3) the combination of these elements.

Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. Cutter Business Technology Journal, 31(2), 47–53.

Trust

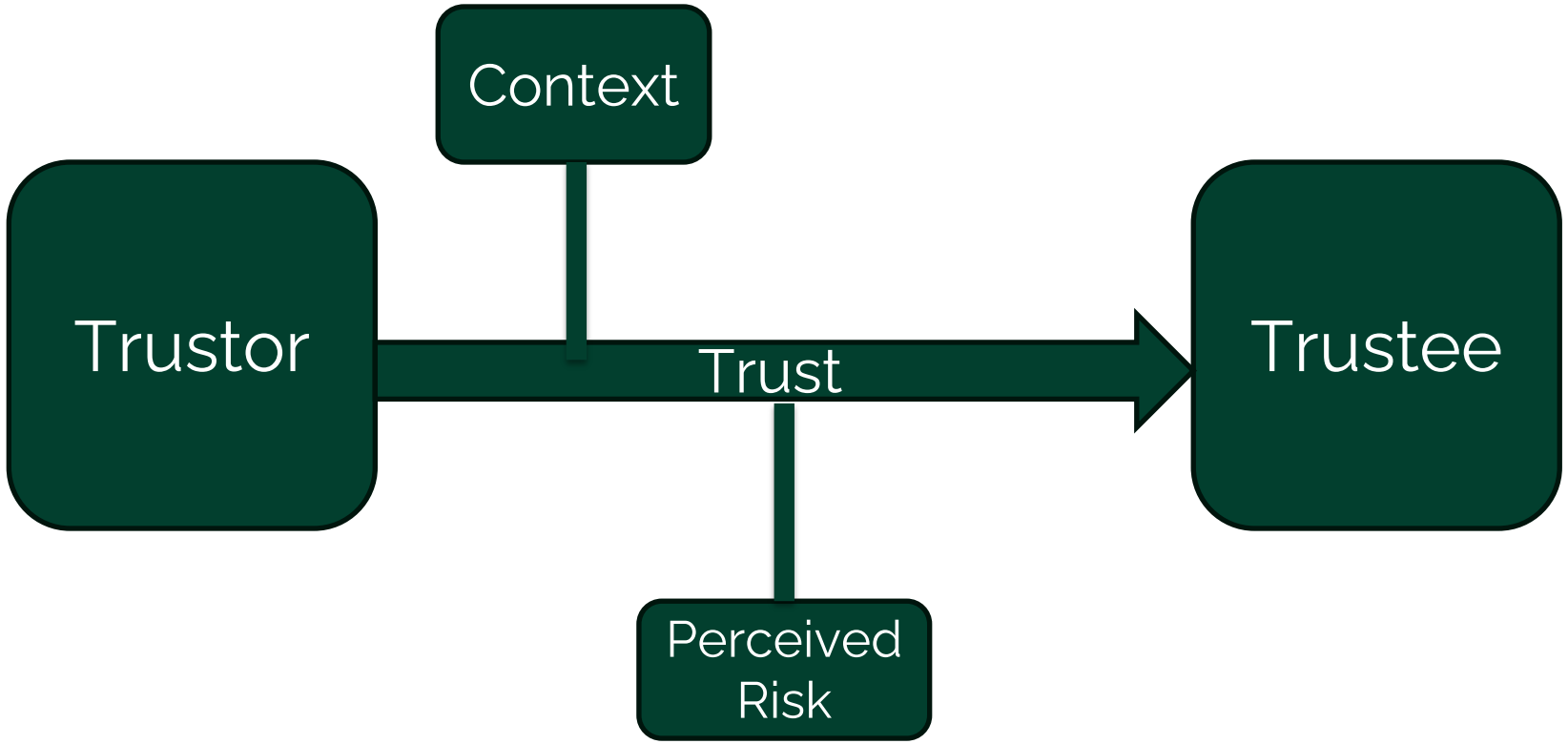
Mayer, R. C. ., Davis, J. H. ., & Schoorman, F. . D. (1995). An Integrative Model of Organizational Trust. Academy of Management Review, 20(3), 709–734.
Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An Integrative Model of Organizational Trust: Past, Present, and Future. Academy of Management Review, 32(2), 344–354.

Trust



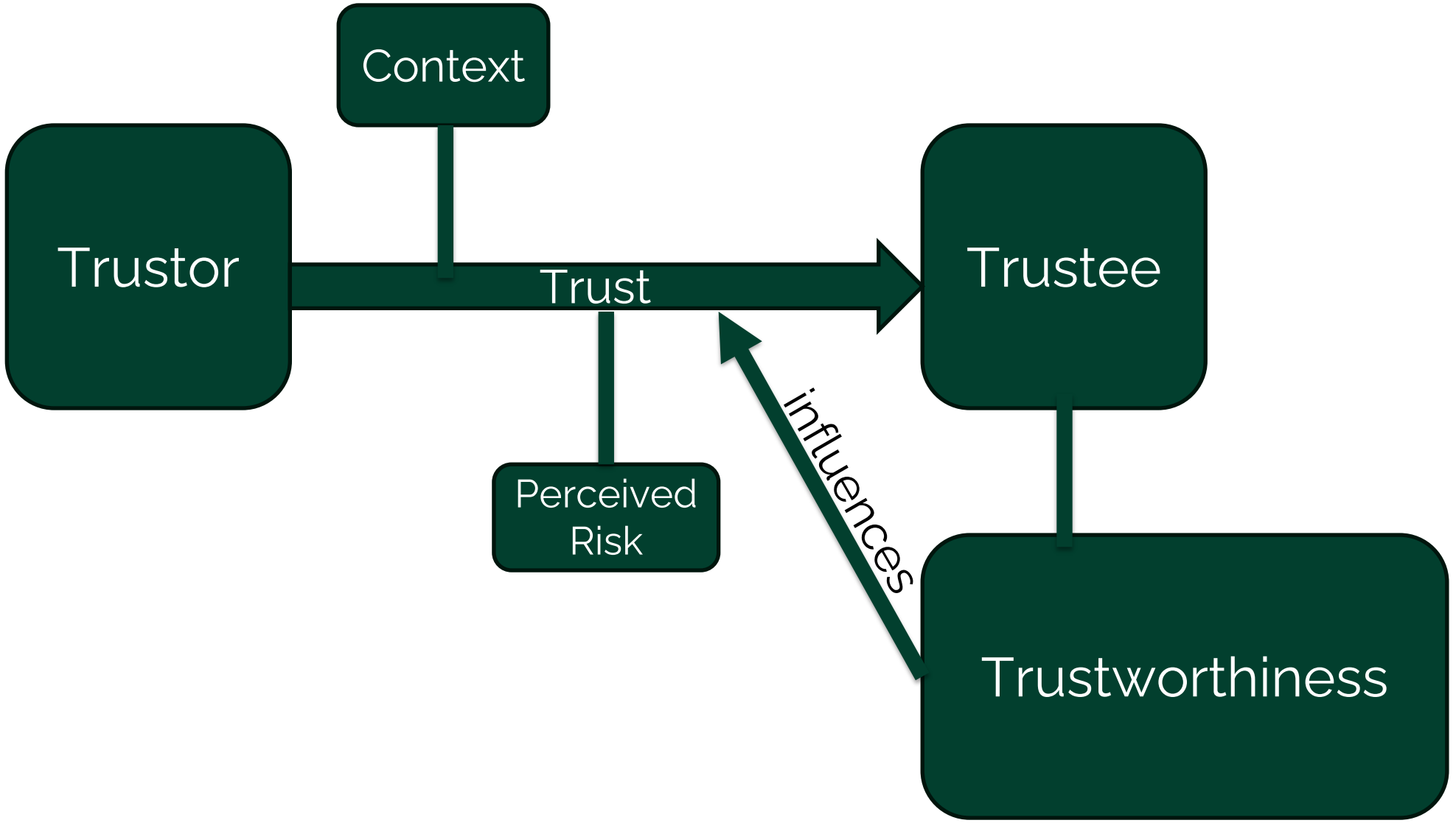
Mayer, R. C. ., Davis, J. H. ., & Schoorman, F. . D. (1995). An Integrative Model of Organizational Trust. Academy of Management Review, 20(3), 709–734.
Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An Integrative Model of Organizational Trust: Past, Present, and Future. Academy of Management Review, 32(2), 344–354.

Trust



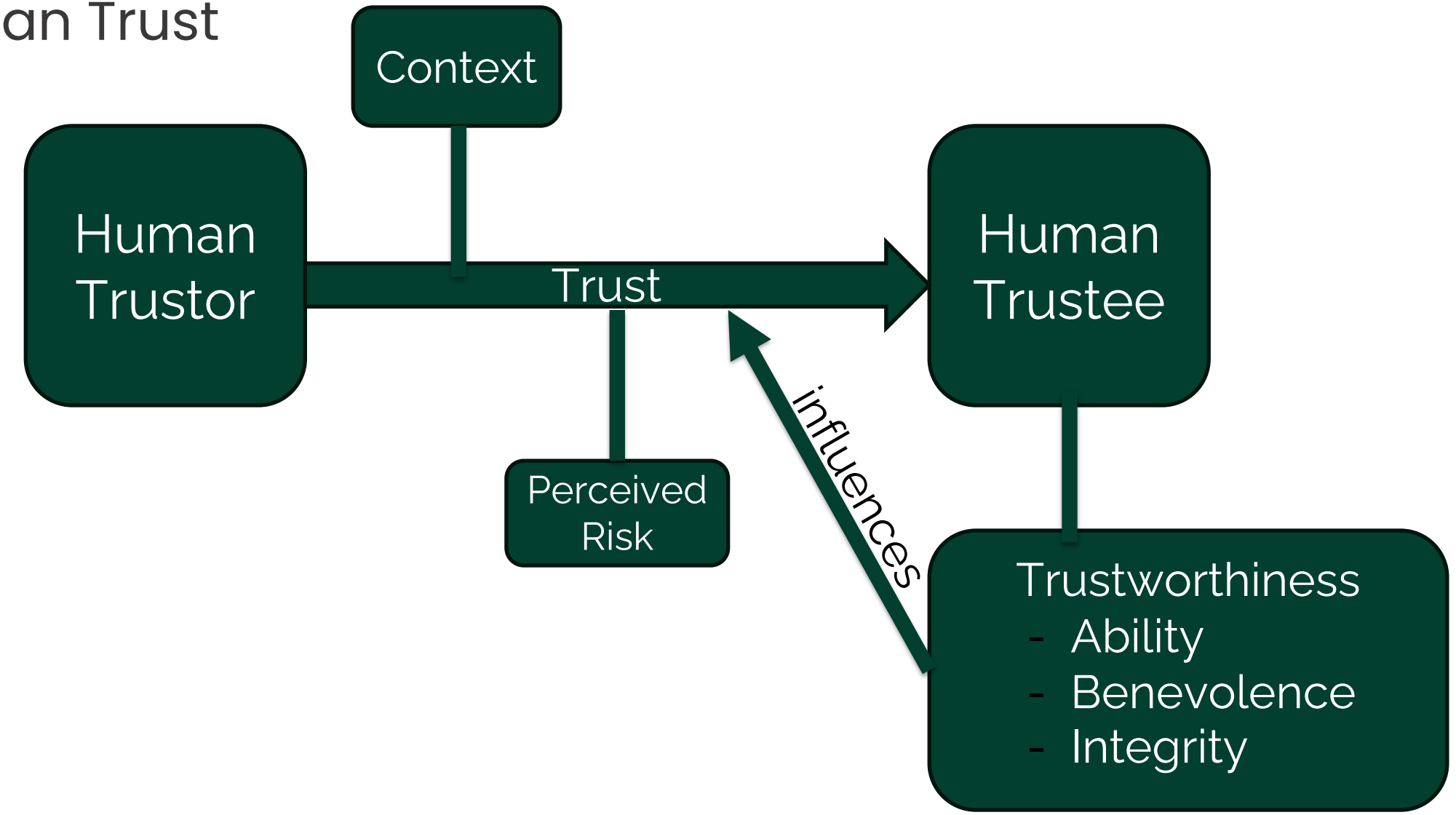
Mayer, R. C. ., Davis, J. H. ., & Schoorman, F. . D. (1995). An Integrative Model of Organizational Trust. Academy of Management Review, 20(3), 709–734.
Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An Integrative Model of Organizational Trust: Past, Present, and Future. Academy of Management Review, 32(2), 344–354.

Trust



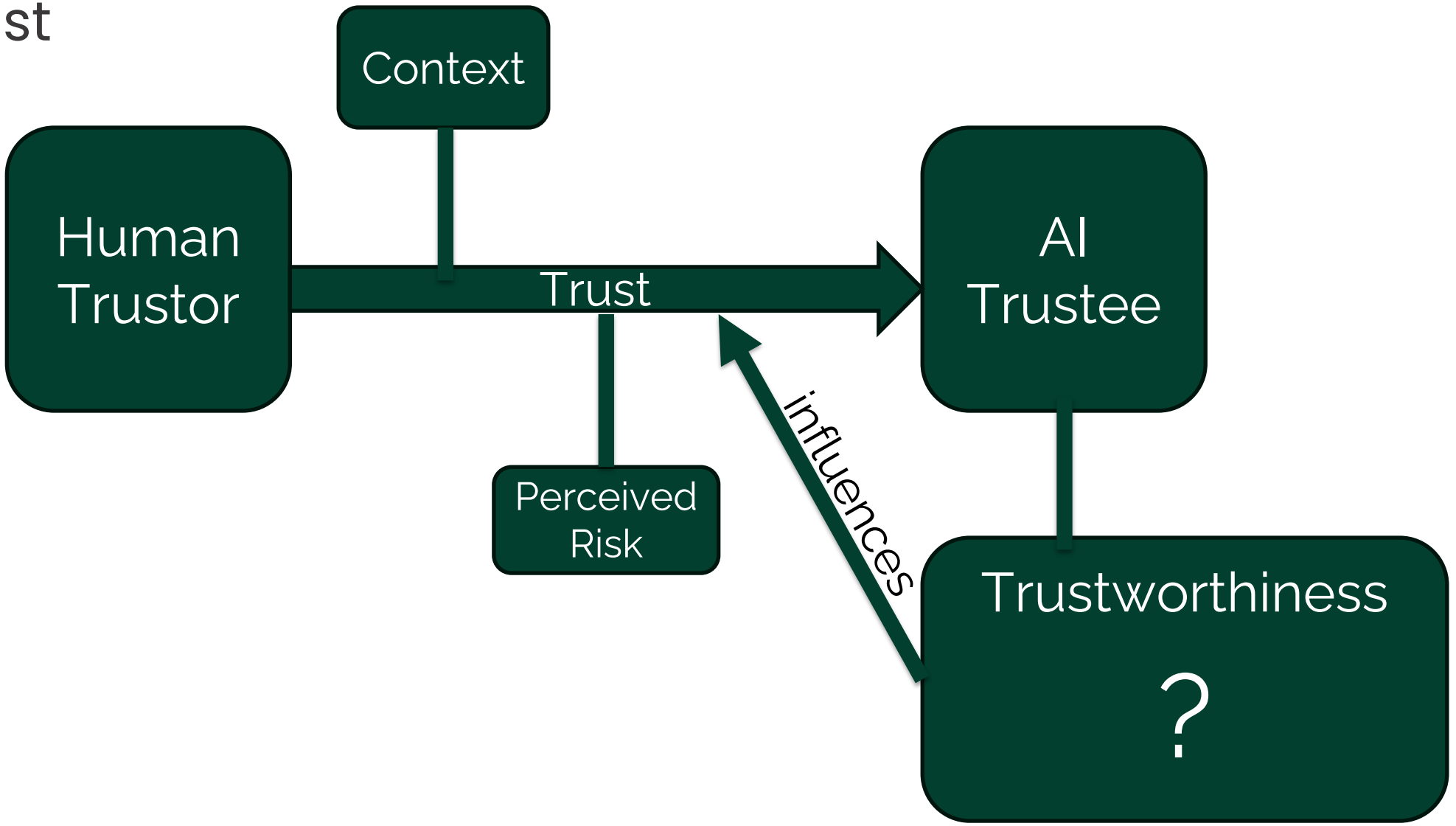
Mayer, R. C. ., Davis, J. H. ., & Schoorman, F. . D. (1995). An Integrative Model of Organizational Trust. Academy of Management Review, 20(3), 709–734.
Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An Integrative Model of Organizational Trust: Past, Present, and Future. Academy of Management Review, 32(2), 344–354.

Human Trust



Mayer, R. C. ., Davis, J. H. ., & Schoorman, F. . D. (1995). An Integrative Model of Organizational Trust. Academy of Management Review, 20(3), 709–734.
Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An Integrative Model of Organizational Trust: Past, Present, and Future. Academy of Management Review, 32(2), 344–354.

AI Trust



Mayer, R. C. ., Davis, J. H. ., & Schoorman, F. . D. (1995). An Integrative Model of Organizational Trust. Academy of Management Review, 20(3), 709–734.
Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An Integrative Model of Organizational Trust: Past, Present, and Future. Academy of Management Review, 32(2), 344–354.

Trustworthy AI

AI HLEG Report

- European Perspective on TAI
- Key Concepts
- Guidelines

European Commission (2019). Ethics Guidelines for Trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Key Requirements of Trustworthy AI

1. Human Agency & Oversight
2. Technical Robustness & Safety
3. Privacy & Data Governance
4. Transparency
5. Diversity, Non-Discrimination & Fairness
6. Environmental & Societal Well-Being
7. Accountability

European Commission (2019). Ethics Guidelines for Trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Part III
Current Research

Perceived Trust in ENFIELD Domains

RQ1 How is trust in the application of AI in the ENFIELD domains perceived by the general public?

RQ2 Which factors influence the perceived trust?

Method Vignette-based Survey

Survey Outline

Demographics & Control Questions

Vignettes

Random

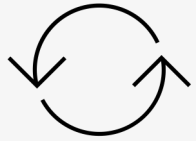
A

OR

B

S-TIAS

Open Questions & Factor Ranking

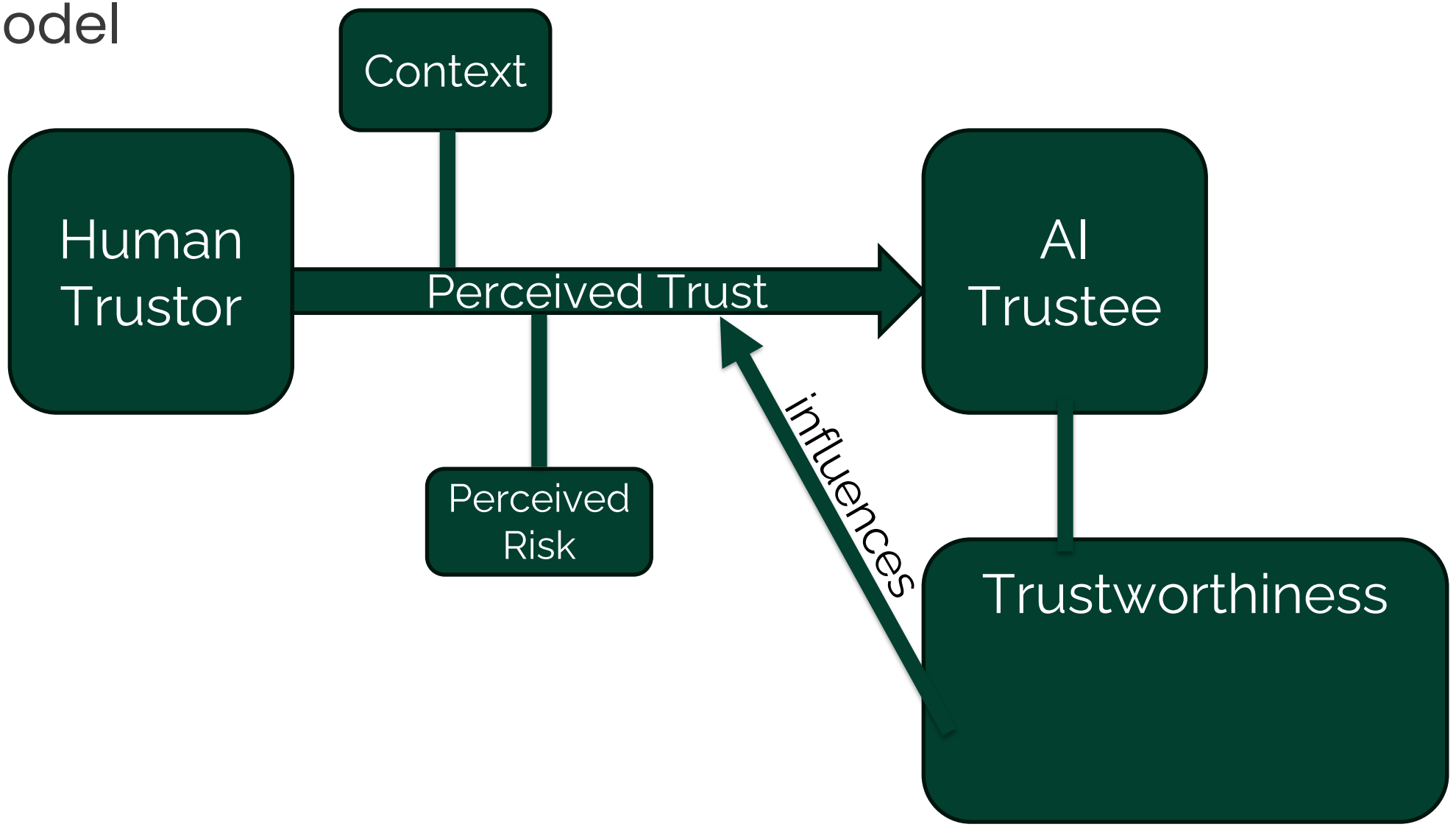


Vignette Sample

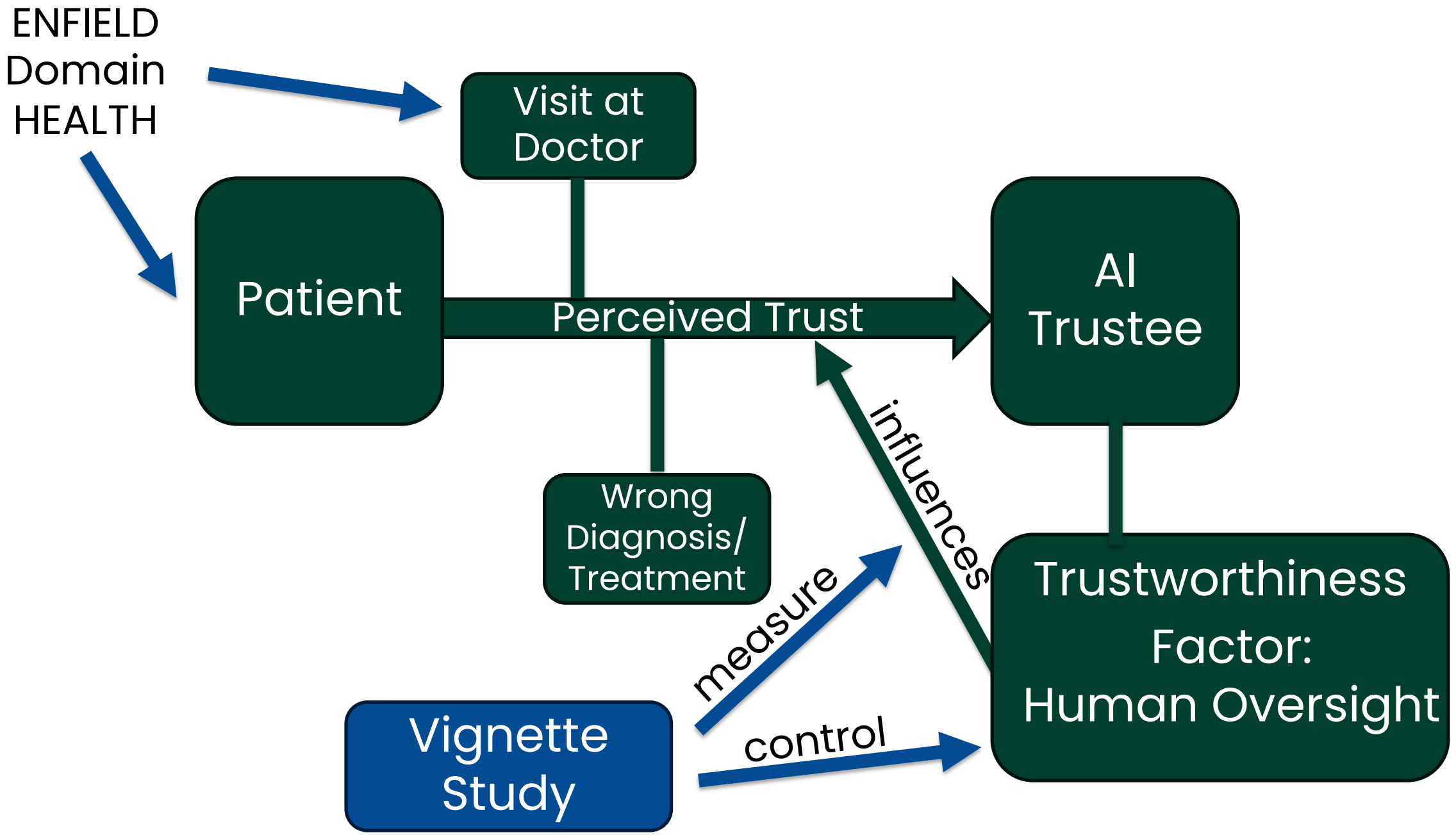
Domain: HEALTH **Factor**: Human Oversight

You are consulting a doctor after experiencing symptoms that concern you. Before your appointment, you learn that the doctor uses an AI system for diagnosis and creating treatment plans. On a regular basis, **[5/10]** % of decisions made by the AI are selected for review by experienced doctors. You would be informed in the event that your diagnosis and treatment have been found to require adaptation. During the consultation, the doctor explains that the AI has analyzed your medical records, lab results, and symptoms. Based on this analysis, the doctor informs you that you have been diagnosed with diabetes.

Trust Model



Mayer, R. C. ., Davis, J. H. ., & Schoorman, F. . D. (1995). An Integrative Model of Organizational Trust. Academy of Management Review, 20(3), 709–734.
Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An Integrative Model of Organizational Trust: Past, Present, and Future.
Academy of Management Review, 32(2), 344–354.


ENFIELD
Domain
HEALTH
Visit at
Doctor
Patient
Perceived Trust
AI
Trustee
Wrong
Diagnosis/
Treatment
influences
measure
Trustworthiness
Factor:
Human Oversight
Vignette
Study
control

Vignette Design Considerations

Domain: SPACE Factor: **Technical robustness and safety** (**Fall-back plans**)

You are about to embark on a long flight and learn that the airline uses an AI-based system to plan the optimal route, specifically designed to avoid areas of turbulence. Weather conditions are monitored in real time to ensure a smooth flight, adjusting the path as necessary to avoid any disruptions. The system has been rigorously tested and **[includes a failsafe system that can take over in case of any technical problems with the AI/can adjust the route as needed to maintain the best possible flight experience]**.

Vignette Design Considerations

Domain: SPACE Factor: Technical robustness and safety (Fall-back plans)

You are about to embark on a long flight and learn that the airline uses an AI-based system to plan the optimal route, specifically designed to avoid areas of turbulence. Weather conditions are monitored in real time to ensure a smooth flight, adjusting the path as necessary to avoid any disruptions. The system has been rigorously tested and [includes a **failsafe system** that can take over in case of any technical problems with the AI/can adjust the route as needed to maintain the best possible flight experience].

Fallback mechanism needs to be technical, not human (confounder human oversight)
Fallback system needs to visibly not improve accuracy (confounder accuracy)
Keep same length and level of detail (confounder: transparency)

Vignette Design Considerations

Domain: SPACE Factor: Technical robustness and safety (Fall-back plans)

You are about to embark on a long flight and learn that the airline uses an AI-based system to plan the optimal route, specifically designed to avoid areas of turbulence. Weather conditions are monitored in real time to ensure a smooth flight, adjusting the path as necessary to avoid any disruptions. The system has been rigorously tested and [includes a **failsafe system that can take over in case of any technical problems with the AI**/can adjust the route as needed to maintain the best possible flight experience].

Fallback mechanism needs to be technical, not human (confounder human oversight)
Fallback system needs to visibly not improve accuracy (confounder accuracy)
Keep same length and level of detail (confounder: transparency)

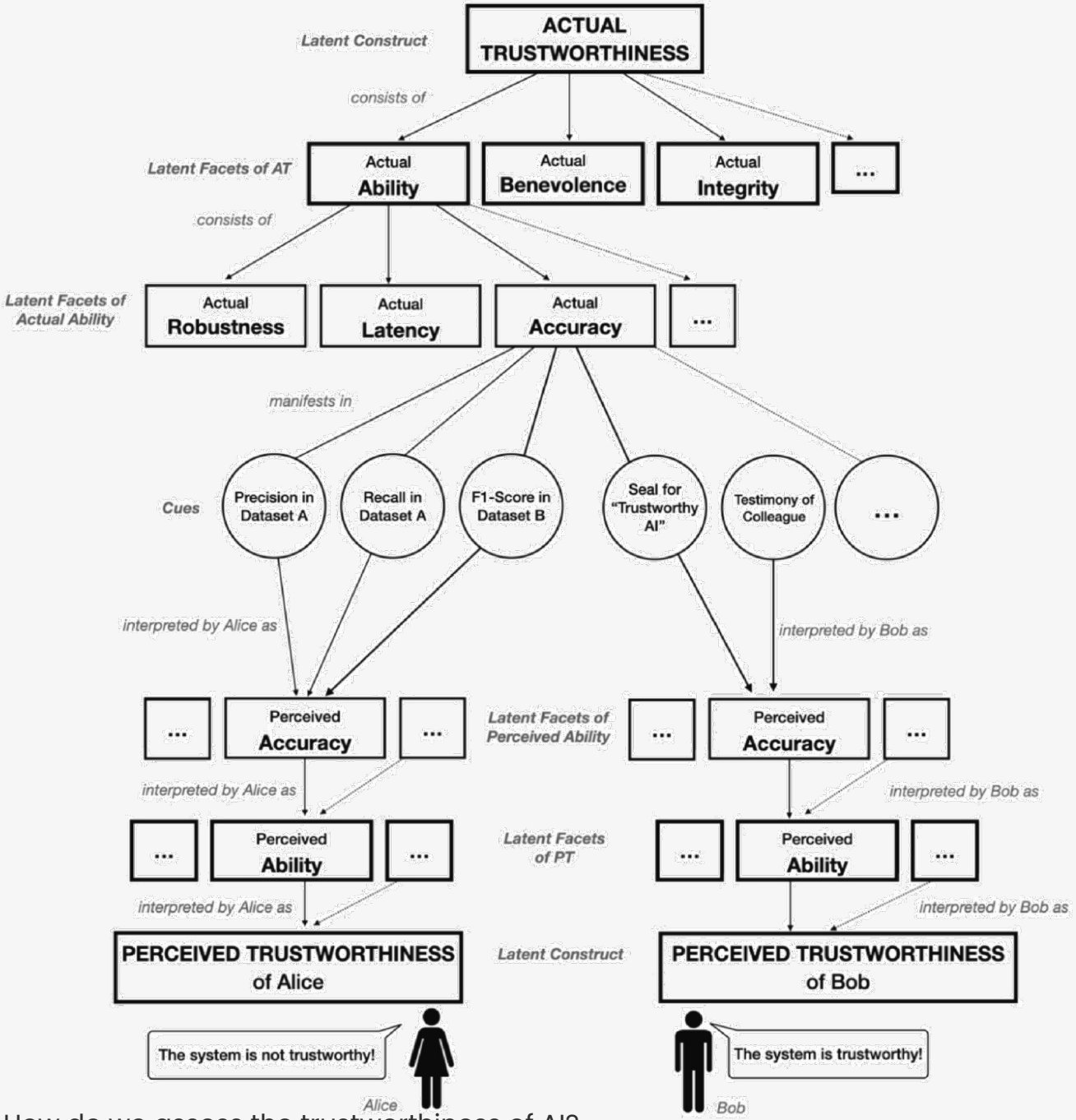
Vignette Design Considerations

Domain: SPACE Factor: Technical robustness and safety (Fall-back plans)

You are about to embark on a long flight and learn that the airline uses an AI-based system to plan the optimal route, specifically designed to avoid areas of turbulence. Weather conditions are monitored in real time to ensure a smooth flight, adjusting the path as necessary to avoid any disruptions. The system has been rigorously tested and [includes a failsafe system that can take over in case of any technical problems with the AI/**can adjust the route as needed to maintain the best possible flight experience**].

Fallback mechanism needs to be technical, not human (confounder human oversight)
Fallback system needs to visibly not improve accuracy (confounder accuracy)
Keep same length and level of detail (confounder transparency)

What are we measuring?



Schlicker, N.et al. (2025). How do we assess the trustworthiness of AI? Introducing the trustworthiness assessment model (TrAM). Computers in Human Behavior, 170, 108671.

S-TIAS: The (Short) Trust in Automation Scale

3 items, 7-level Likert-scaled agreement (1=Not at all, 7=Extremely)

	1 = not at all	2	3	4	5	6	7 = extremely
I am confident in the AI system.	○	○	○	○	○	○	○
The AI system is reliable.	○	○	○	○	○	○	○
I can trust the AI system.	○	○	○	○	○	○	○

Psychometric Validation
S-TIAS is a *valid* and *reliable* measure of human trust in AI

McGrath, M. J., Lack, O., Tisch, J., & Duenser, A. (2025). Measuring trust in artificial intelligence: validation of an established scale and its short form. Frontiers in Artificial Intelligence, 8, 1582880. https://doi.org/10.3389/FRAI.2025.1582880/

Control Scale: Meta AI Literacy Scale

MAILS Short, 10 items, 0-10 rating scale

1. I can tell if I am dealing with an application based on artificial intelligence.
2. I can weigh the consequences of using AI for society.
3. I can use artificial intelligence meaningfully to achieve my goals.
4. I can assess what advantages and disadvantages the use of an artificial intelligence entails.
5. I can program new applications in the field of ”artificial intelligence
6. I can design new AI applications.
7. Although there are often new AI applications, I manage to always be ”up-to date”
8. I can also usually solve strenuous and complicated tasks when working with artificial intelligence well
9. I can handle it when interactions with AI frustrate or frighten me
10. I can prevent an AI from influencing me in my decisions

Carolus, A. et al. (2023). MAILS - Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies. Computers in Human Behavior: Artificial Humans, Volume 1, Issue 2.

Try it out yourself – Participate in our TAI Survey

mytuc.org/hrjd

ENFIELD Summer School 2026

Focus: Trustworthy AI – AI Cybersecurity

Time: July 15-17 2026

Location: Mytilene, Lesbos Island, Greece

Co-located with IPICS (Intensive Programme on Information and Communications Security)

Interested in participating

check

enfield-project.eu/events

Contributing as Speaker

contact me

sebastianheil

Sebastian.Heil@cs.tu-chemnitz.de

Further Resources

- AI HLEG Report “Ethics Guidelines for Trustworthy AI”
- Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., & Zhou, B. (2023). Trustworthy AI: From Principles to Practices. ACM Computing Surveys, 55(9), 1–46. https://doi.org/10.1145/3555803
- Vereschak, O., Bailly, G., & Caramiaux, B. (2021). How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2), 1–39. https://doi.org/10.1145/3476068
- TAILOR Handbook of Trustworthy AI http://tailor.isti.cnr.it/handbookTAI/TAILOR.html

Follow us!

Check our LinkedIn page by scanning the QR Code.

ENFIELD | enfield-project.eu advances adaptive, green, human-centric, and trustworthy AI across Europe. Bringing together academia, industry, SMEs, and the public sector—ENFIELD develops AI solutions for healthcare, energy, manufacturing, and space. By delivering applications, publications, and strategic roadmaps, it fosters reproducibility, ethical deployment, and societal uptake of AI across Europe.

ENFIELD

HorizonEnfield

horizon-enfield

www.enfield-project.eu

Funded by
the European Union

TDW on Trustworthy AI

Theme Development Workshops

March 6th, 2026
Paris, France (Hybrid event)

Thank you!

Funded by the European Union