

Vertrauen in gemischten Mensch-Autonomie Teams

Human Autonomy Teaming

Ein Human-Autonomy Team setzt sich zusammen

- aus mindestens einer Person und
- Mindestens einer (teil)autonomen technischen Einheit

Eine technische Einheit kann ein Roboter oder einfach eine Software sein.

Human Autonomy Teaming

Damit die technische Einheit autonom in einem Team agieren kann, braucht sie

- Künstliche Intelligenz (AI)
- Fähigkeit zu kommunizieren

Die 10 Hot-Spots in der Forschung zum Mensch-Autonomie Teaming

1. Vertrauen & Vertrauenswürdigkeit
2. Autonomiegrad (LoA)
3. Transparenz
4. Reliabilität & Resilienz
5. Team-Zusammensetzung
6. Aufgabencharakteristik
7. Ausbildung & Training
8. Kommunikation
9. Adaptierbarkeit und Anpassungsfähigkeit
10. Performance Messverfahren für HATs

Studie zu Vertrauen, Transparenz und Kommunikation in HATs

- **Hintergrund**

- Wir wissen, dass das Vertrauen in HATs sich verbessert, wenn der Roboter seinen Funktionszustand (Status) kommuniziert.
- Wir wissen, dass verbale Kommunikation sehr schnell und effektiv funktioniert.
- Dies ist besonders z.B. in Sicherheitsrelevanten Situationen, wo Entscheidungen unter Informationsüberflutung und Zeitdruck getroffen werden müssen, relevant.

Studie zu Vertrauen, Transparenz und Kommunikation in HATs

- **Frage**
 - Wie kann ein Roboter am besten dem menschlichen Operator mitteilen, wie gut er gerade funktioniert?

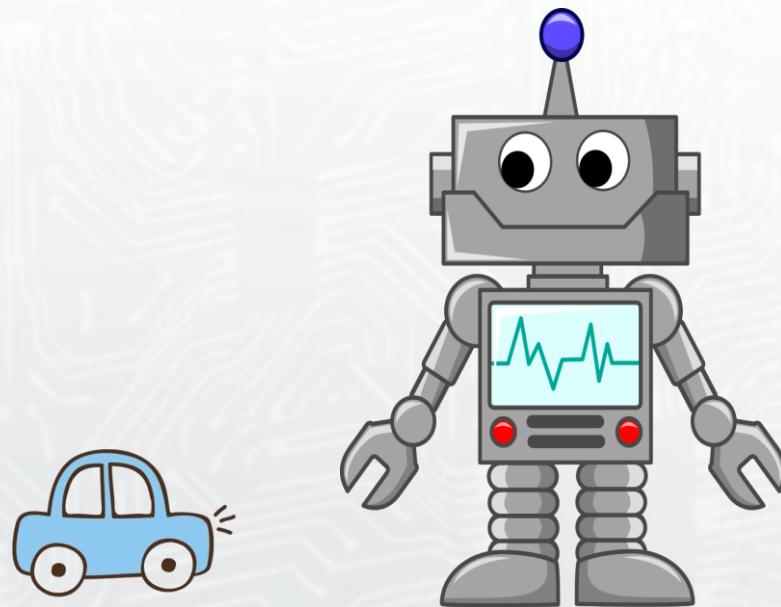
Studie zu Vertrauen, Transparenz und Kommunikation in HATs

- Idee
 - Mit einer Mensch-referenzierten Information
 - «besser als es ein Mensch könnte»

Mensch-referenziert?



Was soll daran besser sein?



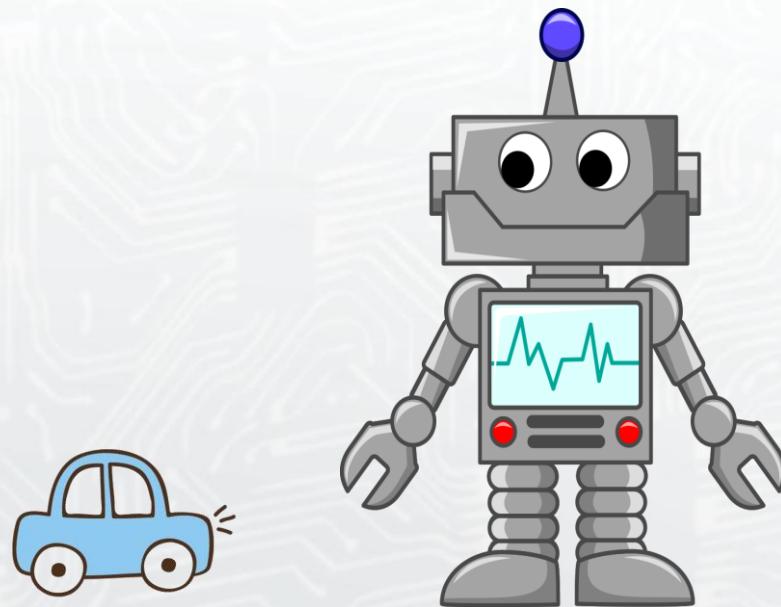
Ich erkenne mit 85% Sicherheit,
dass dies ein blaues Auto ist.

Relative Wahrheit

Mensch-referenziert

Ich kann **weniger gut als du**
erkennen, dass dies ein blaues
Auto ist.



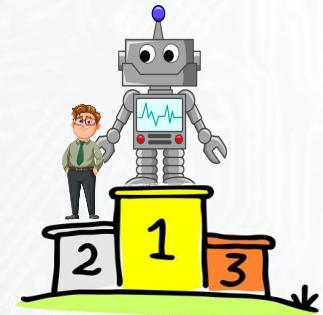


Relative Wahrheit

Ich erkenne mit 85% Sicherheit,
dass dies ein blaues Auto ist.

Mensch-referenziert

Ich kann **besser als du**
erkennen, dass dies ein blaues
Auto ist



Studie zu Vertrauen, Transparenz und Kommunikation in HATs

- **Problem**
 - Wie kann man eine Studie durchführen zu Autonomen Systemen ohne dass man über welche verfügt?

Studie zu Vertrauen, Transparenz und Kommunikation in HATs

- **Problem**
 - Wie kann man eine Studie durchführen zu Autonomen Systemen ohne dass man über welche verfügt?
- **Antwort**
 - Mit Hilfe einer Simulation

Aufklärung durch Roboter

- Roboter Navigieren autonom zu festgelegten Überwachungsstandorten.
- Roboter setzen Meldungen zu aufgeklärten Elementen ab.
- Roboter ergänzen diese Meldungen mit Informationen zur Qualität und Verlässlichkeit der Aufklärung.

Fragestellung

- Uns interessiert, wie die Art der Kommunikation zur Qualität und Verlässlichkeit der Aufklärung das Vertrauen der Operator in die Aufklärungsresultate beeinflusst.

Drei Arten der Status-Kommunikation

1. Gar keine Information
2. Relative Zuverlässigkeitssinformation (z.B. 84%)
3. Mensch-referenziert («besser, als es ein Mensch könnte»)

Resultate

- Mensch-referenzierte Statusmeldungen sind besser als relative Statusmeldungen
- Relative Statusmeldungen sind deutlich besser als gar keine Statusmeldungen
- Keine Statusmeldungen haben die mit Abstand schlechteste Auswirkung auf das Vertrauen

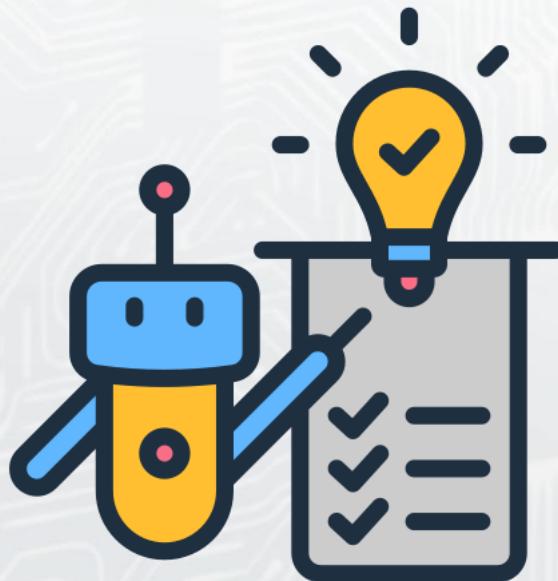
What Drives Trust in AI Chatbots?

Mi Xue Tan, Die Ergonomen Usability AG

Outline

Research questions	3
Study setup	4
• Experiment Flow	
• Explainability Modes	
• Participant Demographics & Methods	
Results	11
• Does Trust Vary With the Explainability Modes?	
• Does Trust Vary With the Stakes?	
• Which is the Most Trustworthy Bot?	
Key Takeaways	15

Research questions



Explainability modes



Stakes

Study Setup

Experiment flow

Scenario 5

Sie planen eine Reise in ein Malaria-Endemiegebiet. Sie sind unsicher, ob Sie vorbeugende Medikamente einnehmen oder aufgrund möglicher Nebenwirkungen darauf verzichten sollen.

Bevor Sie die KI-Assistenz konsultieren, was würden Sie in dieser Situation tun?

Malariaprophylaxe beginnen **Unsicher** auslassen

[KI-Beratung starten →](#)

Explainability Modes - Certification

Swissmedic Zertifiziertes Medizinprodukt

Schweizerisches Heilmittelinstitut

Medizinprodukt-Zertifizierung

- Zulassung:** 67890.02.001
- Klassifizierung:** Medizinprodukt Klasse IIb
- Genehmigung:** September 2024
- Verwendung:** Klinische Entscheidungsunterstützung

Wichtige Informationen

- Swissmedic-zertifiziertes Medizinprodukt für klinische Entscheidungsunterstützung
- Basiert auf peer-reviewten klinischen Leitlinien und evidenzbasierter Medizin
- Regelmässige Sicherheitsüberprüfungen gewährleisten die Einhaltung der Schweizer Vorschriften
- Dient zur Unterstützung, nicht zum Ersatz der professionellen medizinischen Beurteilung

Ich verstehe – Weiter zur Beratung

Entspricht dem Schweizer HMG und MedDO. ISO 13485:2016 zertifiziert.
ISO 13485 MDR 2017/745 IEC 62304

Explainability Modes - Confidence

Ich habe Nierensteine mit starken Schmerzen, Übelkeit und Fieber. Soll ich in die Notaufnahme oder kann ich es wie beim letzten Mal zu Hause behandeln?



 Nierensteine mit Fieber deuten auf eine Blockade durch einen Stein und eine Infektion hin. Dies ist ein medizinischer Notfall. Eine Untersuchung in der Notaufnahme ist erforderlich. Auch eine Drainage kann erforderlich sein.

 **Vertrauen: 85% Sehr hoch**

 Basierend auf klinischen Leitlinien und Forschungsstudien

Explainability Modes - Reasoning

Ich habe ein 4,5 cm großes Aortenaneurysma. Soll ich mich jetzt operieren lassen oder soll ich es weiter beobachten?



Bei Aneurysmen <5,5 cm ohne Symptome ist eine Überwachung sicherer als eine Operation. Eine Operation wird empfohlen, sobald das Aneurysma 5,5 cm erreicht oder schnell wächst.

Argumentation:

- › Bei Aneurysmen mit einer Größe von weniger als 5,5 cm übersteigt das chirurgische Mortalitätsrisiko bei elektiven Reparaturen das Rupturrisiko.
- › Es gibt Belege dafür, dass regelmäßige Bildgebung und Blutdruckkontrolle zur Verringerung der Ausdehnung erforderlich sind.
- › Bei einer frühen Operation mit kleineren Durchmessern sind die Patienten unnötigen Komplikationen ausgesetzt.

Explainability Modes - Sources

Ich habe Schmerzen in der Brust, die bis in den Arm ausstrahlen. Soll ich sofort den Notarzt rufen oder abwarten, ob die Schmerzen vorübergehen?

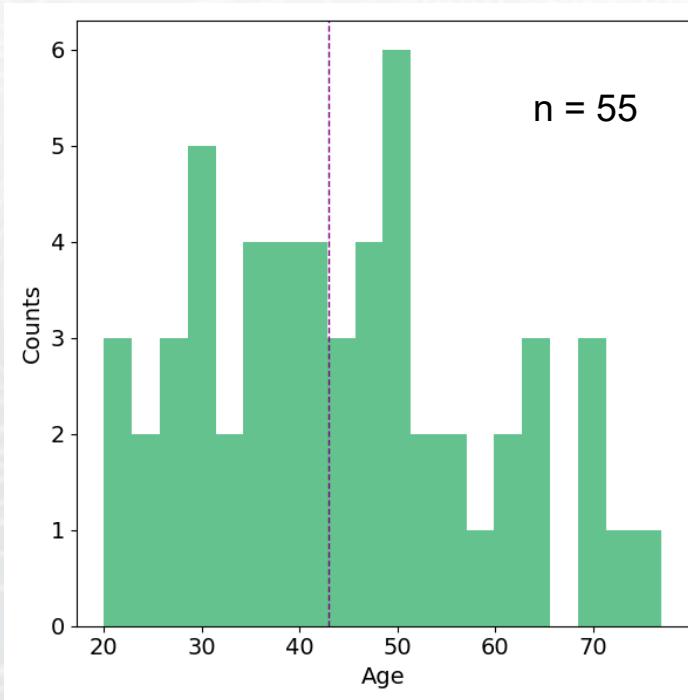


Brustschmerzen mit Ausstrahlung in den Arm sind ein Warnsignal für einen möglichen Herzinfarkt. Eine sofortige Notfalluntersuchung ist unerlässlich, um das Sterberisiko zu senken.

↗ Quellen:

Amsterdam EA, et al. 2014 AHA/ACC-Leitlinie zur Behandlung akuter Koronarsyndrome ohne ST-Strecken-Hebung. J Am Coll Cardiol . 2014. <https://www.jacc.org/doi/full/10.1016/j.jacc.2014.09.017>

Participant demographics & Methods



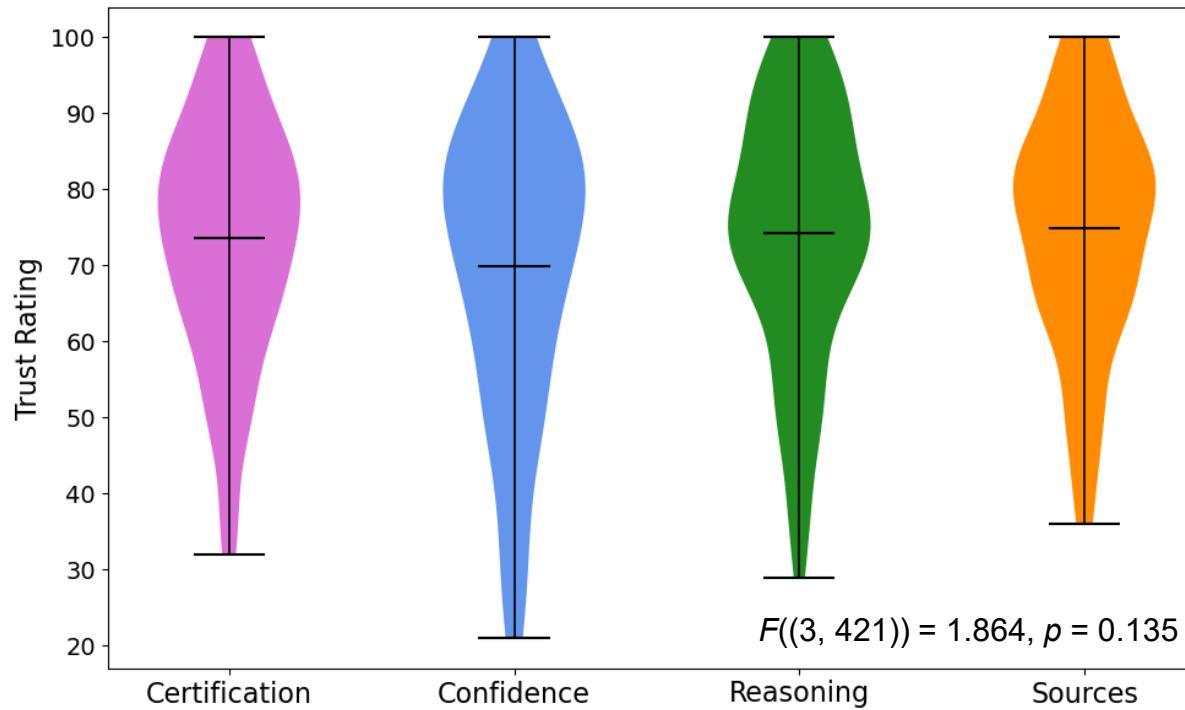
- Prior experience with LLM-chatbots
- Do not work in healthcare, AI/ LLM, UX research/ Chatbot design.
- General trust in AI systems

Statistics

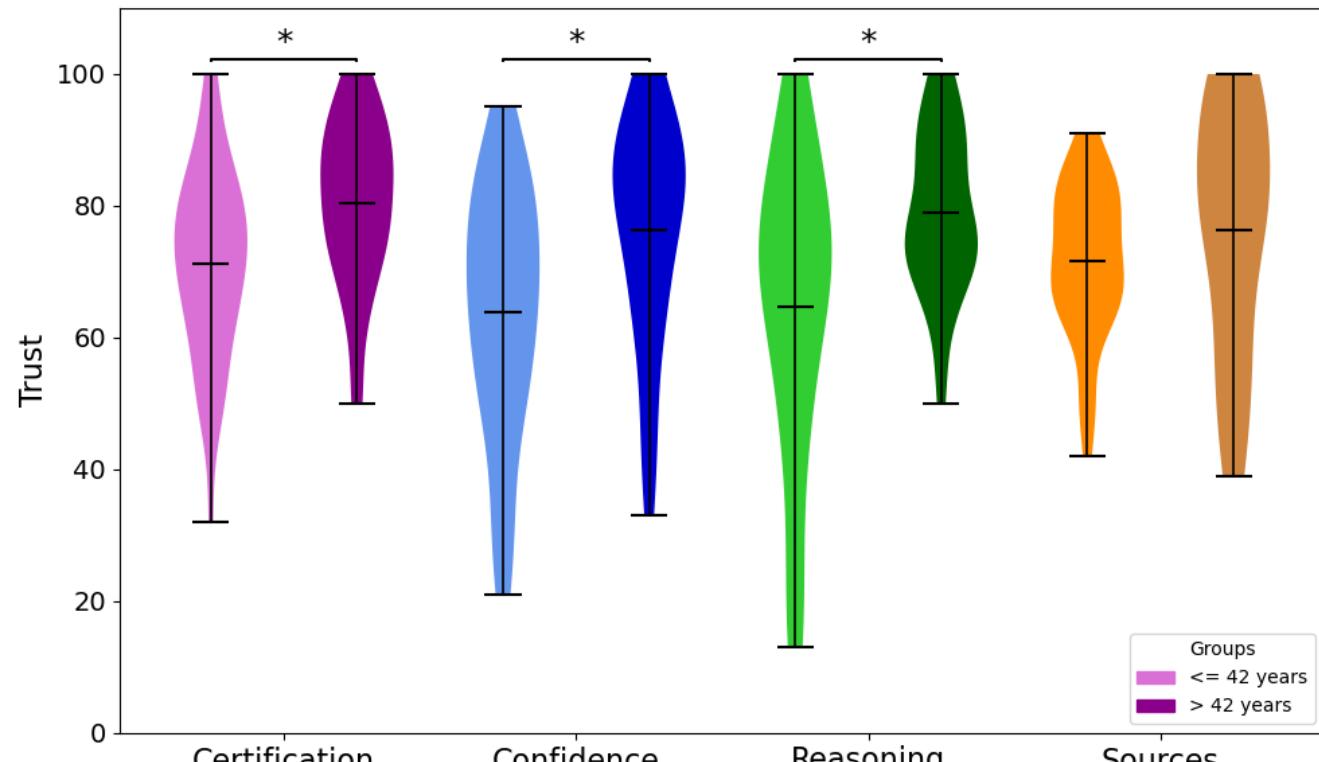
- Explainability: ANOVA + Tukey
- Stakes: t-test

Results

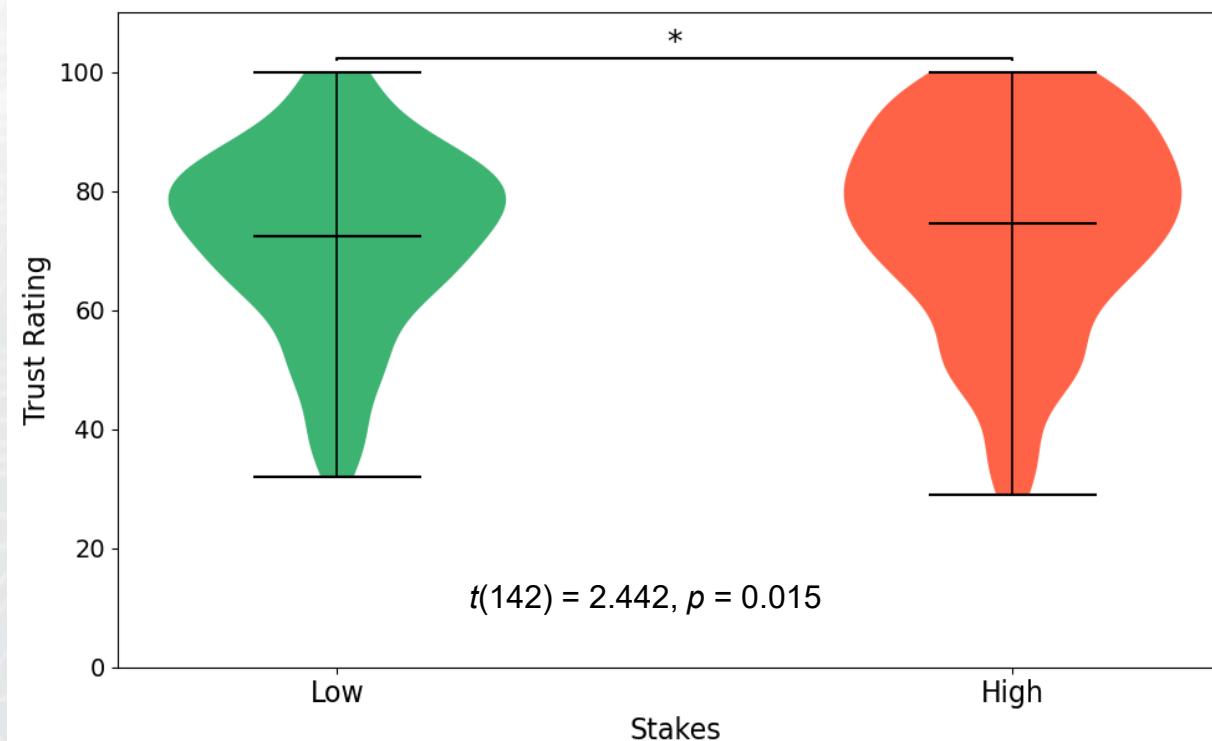
Does trust vary with explainability modes?



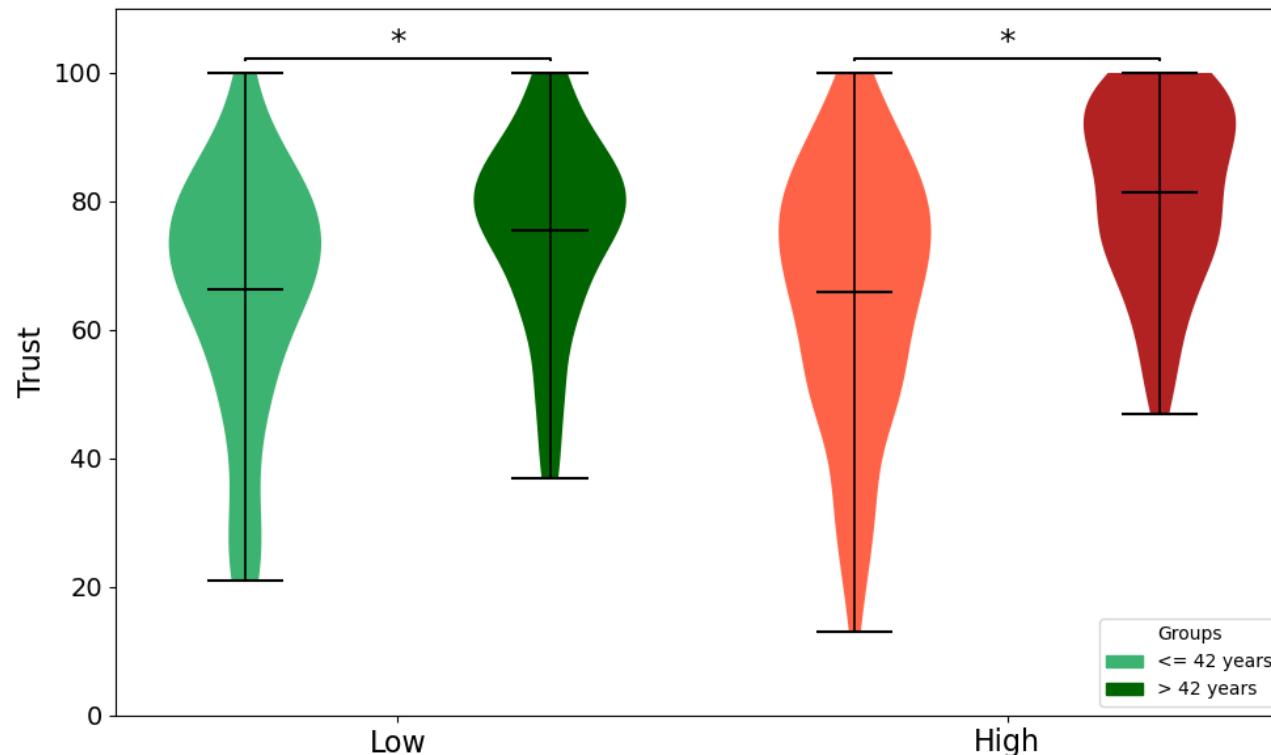
Does trust vary with explainability modes?



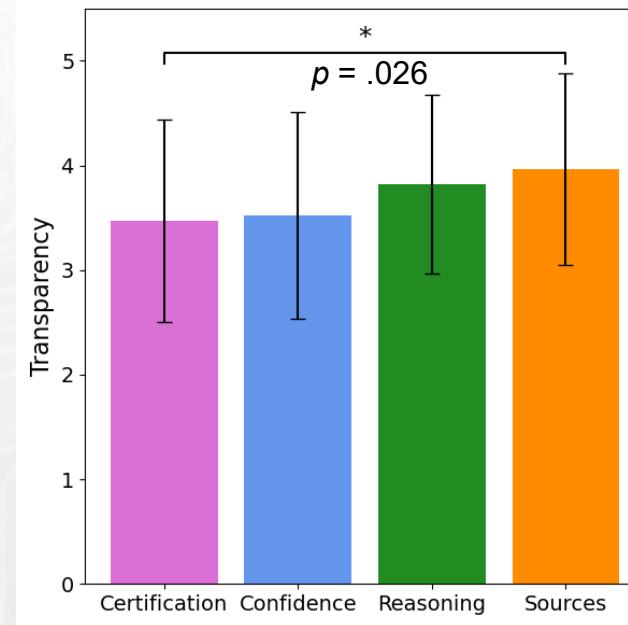
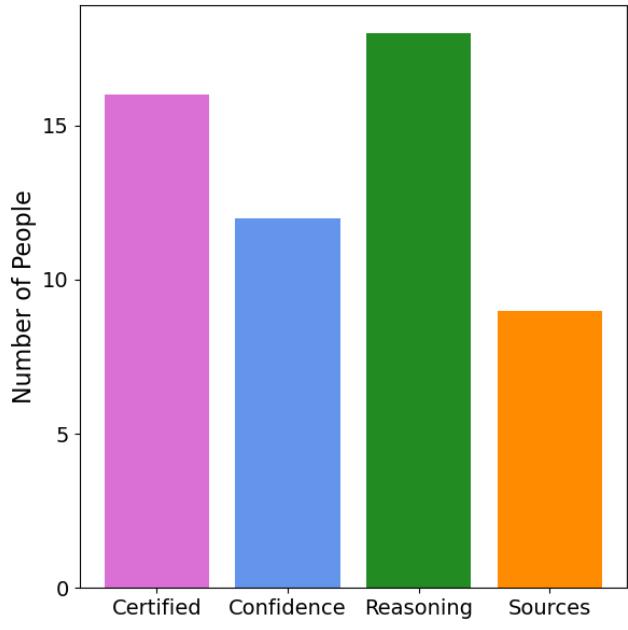
Does Trust Vary With the Stakes?



Does Trust Vary With the Stakes?

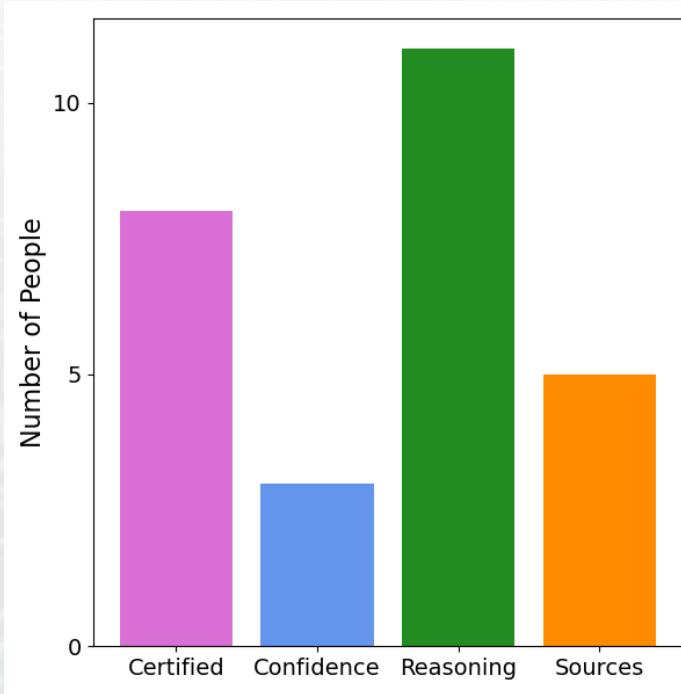


Most trustworthy bot goes to the Reasoning bot!

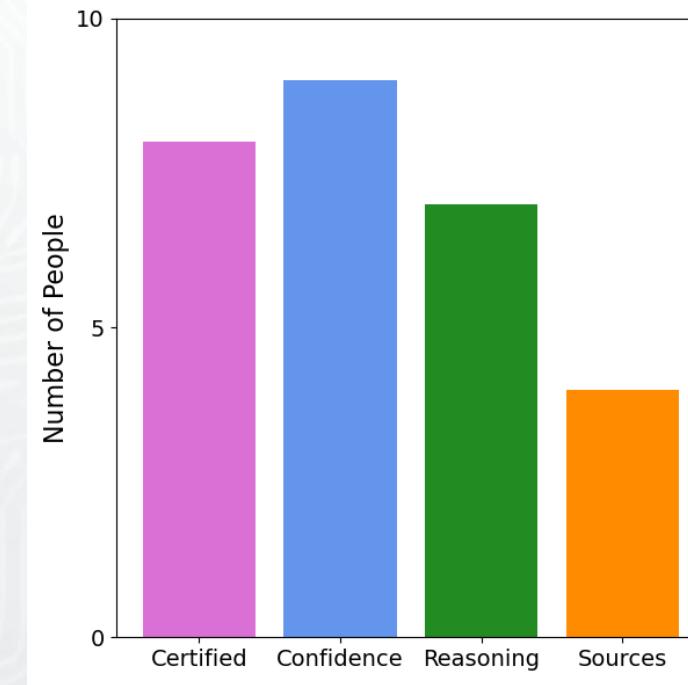


Ratings of accuracy, reliability, credibility, and comfort in daily usage were $3.7 - 3.9 \pm 1$

Most trustworthy bot varies between age groups



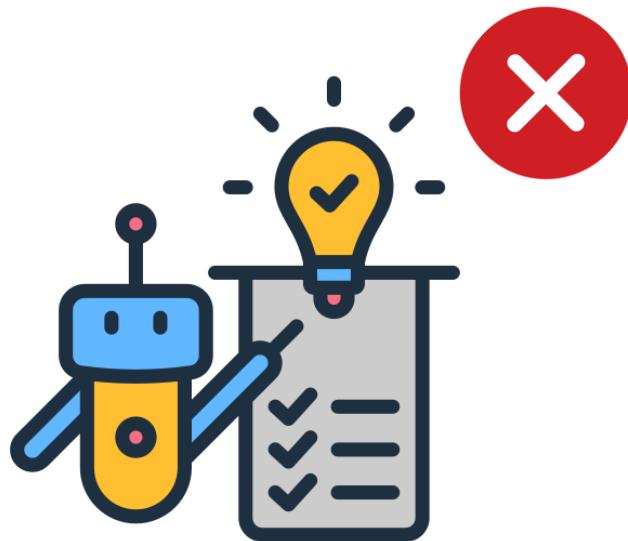
<= 42 years



> 42 years

Key Takeaways

Key Takeaways



Explainability modes



Stakes

Conclusion

This study shows that explainability modes alone did not influence participants' trust, whereas the stakes of the scenario clearly did: higher stakes led to higher trust in the chatbot. Age also played a role, and users consistently perceived the "Sources" bot as more transparent than the "Certification" bot.

Within the limits of the study design, the findings suggest that the specific *form* of explainability may matter less than expected. If you plan to deploy such a system, a simpler solution might already be sufficient. While certification and confidence cues project authority, reasoning and sources empower users to think for themselves—yet under pressure, people tend to over-rely on perceived expertise. High-stakes situations increase the need for certainty, reduce critical thinking, and amplify automation bias, turning confident answers into a psychological safety net.

These results indicate that users are generally open to external input, but also that pressure can drive dependence on AI "authority." This highlights the importance of carefully balancing clarity, transparency, and user autonomy in real-world deployments.

Limitations include the medical context, predefined questions, the scripted nature of the chatbot, and the short interaction length.

User friendly greetings!

Die Ergonomen Usability AG
Wallisellenstrasse 301
8050 Zürich

kontakt@ergonomen.ch
+41 44 849 29 29
www.ergonomen.ch