Insight Titanium Research

Trusting Artificial Intelligence—An Interdisciplinary Approach
Announcement of New Research Project

by Marisa Tschopp and Marc Ruef
April 2019

scip ag
Badenerstrasse 623
8048 Zürich
Switzerland
www.scip.ch

)SCIP(

No Trust, No Use. In the context of AI, there is a critical underlying assumption: "No trust, No Use". Since AI holds great promises (as well as dangers though), tech-companies and AI enthusiasts are especially concerned about how to build trust in AI to foster adoption or usage. Trust seems like the lasting, kind of mysterious, competitive edge.

What is Trust

Without trust, there would be no families, no houses, no markets, no religion, no politics, no rocket science. According to trust researcher Rachel Botsman, trust is the social glue that enables humankind to progress through interaction with each other and the environment, including technology. Trust can be seen as a psychological mechanism to cope with uncertainty and is located somewhere between the known and the unknown.

TRUST CAN BE DEFINED WITHIN THE INDIVIDUAL, A RELATIONSHIP, A SITUATION OR AS A DYNAMIC PROCESS.

In general, trust is a term associated with dynamic and stable properties, with a wide array of research questions and interrelated terms. The focus can be put on rather static attributes of personal trust factors or on a more dynamic perspective, where trust is perceived as a process – from building, maintaining, and developing trust as well as destroying and regaining trust. Situational factors from time to culture or organizational structures, have to be kept in mind, when exploring moderating effects. A popular, rather market-driven focus is the exploration of trust building conditions (so called antecedents of trust) as well as characteristics of one-on-one trust relationships.

Trust and AI

In the AI context, there are various reasons, why trust has become a very popular research topic. For example, when lives can be saved in the military context, where robots are designed to work with or replace humans in high-risk situations (Yagoda et al. 2012). Or in everyday life, when it comes to self-driving vehicles. Trust has always played a big role in the adoption of new markets, technologies or products, like e-commerce, the fax or mobile phones. It will continue its impact on business and economic behavior, for instance, if you consider the use of digital assistants like Alexa to shop for food.

FACTORS FOR TRUSTING AI : PERFORMANCE, PROCESS, PURPOSE.

So far, research has agreed upon the similar three main technological characteristics of trust.

- Performance: Does it perform well? Is it safe? Is it built correctly?
- Process: Does it perform the way we intended? Can we predict the outcome?
- Purpose: Do I have a good feeling about the intent of the program and the provider? Does it adhere to ethical standards? Is the AI trustworthy?

Research Question

Human trust has been examined in psychology for a long time now, however, within the context of Artificial Intelligence there are many more open research questions. The first natural step that comes to mind, is to focus on the capabilities of a product. That is why our research department developed the AIQ, a method to measure competences of conversational AI, to build trust in form of a proof of quality signal. The same intent applies to endeavors of building transparent or explainable AI, who aim to build trust through trustworthiness as feature of the product or image. However, as the market is still in an early stage, this may be too fast. Our research is taking a step back and focuses on the less obvious factors that influence trust building in the context of AI, especially from the perspective of the general public.

WHAT DOES IT TAKE TO TRUST AI?

With this study we want to 1) Explore trust and distrust in general, 2) Analyze the factors, that contribute to the initial formation of trust and distrust in Artificial Intelligence, and 3) Investigate the „no trust, no use“ assumption.

Method

After organizing the Trusting AI sub-track at the Applied Machine Learning Days at EPFL 2019, a questionnaire was created to measure trust-building factors in the context of AI. The questionnaire was reviewed by two independent survey design experts (Germany) and five experts from the tech-industry (Switzerland). It is an explorative study focusing on antecedents of trust, attitude and personality factors (Big 5 scale) using quantitative and qualitative methods. It is targeted towards the general public and aims at 100 participants.

Significance

This study wants to approach trust and AI from an interdisciplinary perspective, focusing on antecedents of trust and the trust-usage relationship. It shall contribute in three ways: First, we examine theories and knowledge about trust and AI across disciplines in academia as well as in practice. Hence this study aims to expand knowledge in applied psychology, human-machine-interaction, and technology management. Secondly, our results can be useful in various ways: For developers and from a management perspective. For academics in interdisciplinary applied fields, to generate new hypotheses. Last but not least, it is useful for the general public, having the opportunity to engage with this critical issue on a low threshold by taking part in this study.

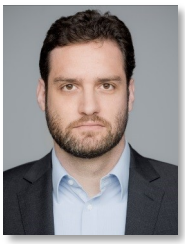
References and selected readings

[1] On Trust in AI—A Systemic Approach I M. Tschopp I scip AG Blog 2018
[2] Who Can You Trust? How Technology Brought Us Together and Why It Drives Us Apart I R. Botsman I Book 2017
[3] Artificial Intelligence: The Compendium I M. Ruef et al. I ebook 2018
[4] Selected papers:

- You Want Me to Trust a ROBOT? The Development of a Human-Robot Interaction Trust Scale I R.E. Yagoda et al. I 2012
- A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems I K. Schaefer et al. I 2016
- Applied Artificial Intelligence and Trust—The Case of Autonomous Vehicles and Medical Assistance Devices I Hengstler, M. et al. I 2016

Link to survey: https://forms.gle/KMbdXtYUaA9RgVPF9

About the Authors



Marc Ruef is Head of Research and Member of the Board at scip ag. His focus is on conducting research in the broad field of cybersecurity, including computer and network security, vulnerability assessment and penetration testing. As an entrepreneur and expert in information technology he has extensive expertise in the development, application and capabilities of artificial intelligence (AI) solutions with specific interest on language and reasoning capabilities, taking security related, business and societal implications into account. Marc has published various best-selling books, more than 400 white papers and media articles, is teaching at several higher education institutions and is a frequent speaker at conferences and workshops in the public and private sector worldwide.

Marisa Tschopp is a researcher at scip ag and ambassador for Women in AI Switzerland. Her focus is on conducting research about AI-based systems from a humanities perspective, with a wide range of questions related to psychological phenomena, governance and ethical implications. As an organizational psychologist she has experience in social and educational institutions with specific passion for digital teaching-learning trends, and has been teaching in several higher education institutions in Germany and Switzerland. She published various media articles, books, and papers and is a frequent speaker on conferences and events worldwide. Marisa holds a Master's degree in Psychology of Excellence in Business and Education from the Ludwig-Maximilians-University of Munich, Germany as well as a BA business degree, focusing on market and consumer psychology.

Mission: Know the Future

The Titanium Research Team is the highly independent for-profit research department which explores essential issues and innovations in the field of emerging technologies. We pursue an interdisciplinary approach comprising technical and non-technical considerations to guarantee the preparation of our customers in the best possible way.

Our research endeavors include a range of projects tapping into various activities such as knowledge building, education, training, consulting and technical & non-technical development. We want to build knowledge, strengthen collaboration between research, industry, policy makers and users in order to explore and comprehend the nature of digitalization and emerging technologies.

scip ag is a Swiss company providing consulting services in the field of cybersecurity. Security is our business. https://www.scip.ch/en/