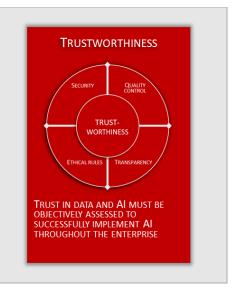# TRUSTWORTHINESS

Trust in data and AI must be objectively assessed to successfully implement AI throughout the enterprise. The coupling of humans and AI enhances the intelligence of the organization, provided they complement each other to ensure reliable management. To achieve this, the user's trust in AI must be strong and can be improved by promoting AI that upholds the following qualities: reliability, honesty, competence, and integrity.



## 1. CONDITIONS OF SUCCESS

With generative, symbolic, or analytical AI, the dialogue between humans and computers is not limited to the deterministic scope of traditional software. Indeed, AI adapts to management situations by considering unforeseen events and incomplete information. Thus, the user no longer merely manages data to execute a predefined process but engages in a constructive dialogue with the AI to obtain responses tailored to their work situation.

For example, when a doctor classifies domestic accidents according to administrative criteria, they select values in the management application's interface: the time slot of the accident, location, object involved, height of the fall, water level, type of fire, etc. This data is used for statistical studies. The more precise the classification, the more time this administrative task consumes for the doctor. With AI, it is no longer necessary to predefine possible classifications in advance. The practitioner simply expresses the accident's context in natural language, and the AI handles its classification. By using voice input processed by the AI, the doctor further reduces the time spent on classifying each accident. The old application, at least its user interface, becomes obsolete. As such, the scope of digitization through AI is broader than that of traditional software.

TRAIDA advises first leveraging this strength to enhance productivity without changing existing applications and processes, and then focusing on creativity to deeply transform the organization and applications (see TRAIDA's business cards on productivity and creativity).

In other words, AI invites the user to contribute knowledge, clearly articulate their requests, analyze the responses, and ask for clarifications or additional information when needed. Thanks to this more intelligent dialogue between humans and machines, new task automations become possible. This setup is especially powerful for logics not fixed in algorithms, benefiting from the collaboration between the user and the machine.

**The need for trust**

This human-AI coupling increases the organization's intelligence, provided they complement each other to ensure reliable management. To achieve this, the user's trust in the AI must be strong, built on the following qualities:

- **Reliability**: Working professionally within a domain of expertise.

- **Honesty**: Telling the truth and not seeking to deceive or manipulate others.

- **Competence**: Possessing the necessary skills to be reliable in the domain of expertise.

- **Integrity**: Acting according to ethical standards and the company's values.

These qualities form the foundation of trust required to successfully deploy AI at an enterprise scale.

## AI and consciousness

At the time of writing this card (October 2024), AI does not possess consciousness. It does not, therefore, conceptualize its relationship with humans. However, the data used to train AI essentially injects a cognitive heritage into it, embodying a form of artificial consciousness. Indeed, this heritage contains deliberate biases and others introduced by error. Thus, depending on its learning process, an AI is more or less confident in the role for which it was trained, such as pattern recognition, data analysis in a domain of expertise, solving specific problems, etc. **To avoid usage errors, it is important that AI refuses to respond to requests for which its training is insufficient**. If this boundary does not exist, AI responds outside its field of competence with hallucinations and approximations, which deteriorate the user's trust. Conversely, when this boundary is in place, a form of artificial consciousness emerges.

Thus, the decision of AI to respond or not to a user's request is a first level of singularity. It shifts AI into a realm beyond just new technology. With generative AI, the degree of tolerated hallucination is a parameter that adjusts the expected level of trust in the responses. The more the user expects a creative AI, the more hallucinations are encouraged; conversely, for a scientific AI that relies on facts, hallucinations will be minimized. Between these two behaviors, the degree of hallucination varies to allow AI to innovate from real facts. Thus, a subtle relationship takes place between the AI user (a), the AI itself (b), and indirectly the system that trained it (c). This relationship becomes systemic when several AIs are deployed within an organization for different users and with distinct training modes. This triplet (a, b, c) multiplies and generates exponential complexity in the interactions, with varying degrees of hallucinations desired or endured.

For example, an AI assistant (a1) manages logistical flows and interacts with another AI assistant (a2) specializing in comparative supplier analysis. A request is made by a1 for the urgent selection of a service provider to solve a delivery issue. However, a2 has not been trained to guarantee reliable contracting with a service provider, but only to establish a comparative list of potential suppliers. AI a2 then responds to AI a1 that it is not capable of producing the requested work. It justifies this decision by reminding that its training scope covers a different need. At this point, AI a1 has two possibilities: either it no longer solicits AI a2 because the risk of contracting in real-time with a service provider is too high; or it forces AI a2 to provide a limited selection so that it can assign an urgent order to a service provider. This use case shows that regulating exchanges between AIs and with humans is not trivial. These exchanges fall outside the usual scope of traditional software programming. Since it is impossible to anticipate all scenarios, it is necessary to build barriers to ensure that AIs do not exceed their area of competence and responsibility.

**These barriers rely on establishing a sufficient level of trust and elevate AI to the rank of a stakeholder within the organization.**

The level of trust is quickly established when it comes to integrating new technology into a company: either it brings a benefit, and trust is built, or it fails, and the company can move on without it. With AI, it is impossible to accept a situation where its usage fails and for the company to simply discard it without facing significant consequences. Indeed, it is human behavior toward AI that can lead to failure, not the technology itself. **In other words, the level of intelligence brought by AI is already too important for it to be disqualified (as of October 2024)**. The company then has no other option but to succeed in its transformation with AI, ensuring a sufficient level of trust in its use.

In TRAIDA's business card on productivity, we explained that AI is more than just a simple technology; it is a new stakeholder that possesses superior intelligence to humans in many use cases. Therefore, an efficient and harmonious relationship between humans and AI requires defining a stable and clear framework for work methods. It is not about claiming perfection in all operational rules from the first deployments of AI but ensuring that this framework is built with respect to the interests of all parties, in a transparent and committed manner. In other words, with an affirmed, durable, and proven level of trust.

This approach calls for resources that are not economic but human. This is a fundamental point for the success of AI deployment at the company-wide level. Indeed, although it is always possible to force an organization to adopt a new technology by imposing work processes, this does not work with AI. It is not enough to implement processes that include new technology; it is necessary to continuously reinvent the relationship between humans and AI. **This is a perpetual re-engineering that requires a critical mindset from all actors in how they collaborate with AI**. Without trust, this perpetual re-engineering is devastating for the organization, and AI risks entirely replacing humans.

However, trust cannot be decreed. It is a quality that is built progressively and is never definitively acquired. The company must then establish an organization that nurtures this trust. TRAIDA advises setting up two independent bodies that address the following objectives:

1. **AI Compliance**: Define the rules for AI transparency, ethics, and security.

2. **AI Quality Control**: Ensure that these rules are applied in accordance with expectations.

Each body operates with its own resources to guarantee its autonomy and independence. They are under the responsibility of the enterprise governance (see TRAIDA's card on this topic).

## 2. IMPORTANCE OF THIS CARD FOR YOUR TRANSFORMATIVE AI

In the first part of this card, we emphasized the importance of trust in AI to ensure its successful large-scale deployment within the company. Building this trust is based on the consideration of three cardinal values: transparency, ethics, and security. To ensure these values are upheld over time, an AI quality control process is added.

### TRANSPARENCY

Transparency is the responsibility of the AI compliance body (see the first part of this card). This involves documenting practices and making them known. The areas of application are varied, such as:

- The uses and impact on employability.

- The data used for training both internally and with stakeholders.

- Ethical and security rules.

- Expected and actual results.

- Investments.

- Risks and opportunities.

- Training and career plans with AI.

- Traceability and auditability of results.

- Detection of deviations and fraud in the use of AI.

- ../..

## ETHICAL RULES

Best practices for AI ethics can be found in the public domain, particularly in government regulations. They address at least the following topics:

- Respect for the company's values.

- Respect for HR policies.

- Compliance with regulations.

- Democratization of AI usage.

- Reduction of the carbon footprint.

The definition of ethical rules is the responsibility of the AI compliance body (see the first part of this card).

## SECURITY

A security breach involving AI inevitably leads to a lack of user trust in the technology. It is therefore important to address this by reviewing all processes that incorporate AI, such as:

- Data protection.

- Licensing for commercial and open-source AIs.

- Onboarding a new employee (AI usage rights, access to training data, etc.).

- Offboarding an employee.

- Work with contractors.

- Mergers and acquisitions.

- Rollback in case of malfunction.

- Backup and archiving.

- Etc.

In general, the stronger the security rules, the fewer innovative uses of the technology are possible. To avoid ossifying practices, it is useful to set up a free-use mode for AI in a technical environment and with data that does not pose security risks. This acts as a kind of sandbox for AI, where users can install new tools, test innovative AI behaviors with fictitious data, and more.

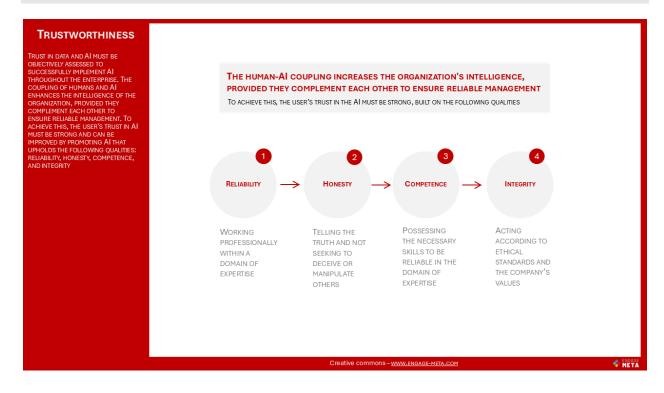The definition of security rules is the responsibility of the AI compliance body (see the first part of this card).

## CONTROL QUALITY

A dedicated AI quality control body is planned within the organization, alongside the body responsible for defining transparency, ethics, and security rules.

This control is applied at all levels of AI involvement: design, budgetary decisions, implementation, training, evaluation of results, etc.

## 3. BLUEPRINT

### TRUSTWORTHINESS

TRUST IN DATA AND AI MUST BE OBJECTIVELY ASSESSED TO SUCCESSFULLY IMPLEMENT AI THROUGHOUT THE ENTERPRISE. THE COUPLING OF HUMANS AND AI ENHANCES THE INTELLIGENCE OF THE ORGANIZATION, PROVIDED THEY COMPLEMENT EACH OTHER TO ENSURE RELIABLE MANAGEMENT. TO ACHIEVE THIS, THE USER'S TRUST IN AI MUST BE STRONG AND CAN BE IMPROVED BY PROMOTING AI THAT UPHOLDS THE FOLLOWING QUALITIES: RELIABILITY, HONESTY, COMPETENCE, AND INTEGRITY

**THE HUMAN-AI COUPLING INCREASES THE ORGANIZATION'S INTELLIGENCE, PROVIDED THEY COMPLEMENT EACH OTHER TO ENSURE RELIABLE MANAGEMENT**
TO ACHIEVE THIS, THE USER'S TRUST IN THE AI MUST BE STRONG, BUILT ON THE FOLLOWING QUALITIES

**1 RELIABILITY** → **2 HONESTY** → **3 COMPETENCE** → **4 INTEGRITY**

WORKING PROFESSIONALLY WITHIN A DOMAIN OF EXPERTISE

TELLING THE TRUTH AND NOT SEEKING TO DECEIVE OR MANIPULATE OTHERS

POSSESSING THE NECESSARY SKILLS TO BE RELIABLE IN THE DOMAIN OF EXPERTISE

ACTING ACCORDING TO ETHICAL STANDARDS AND THE COMPANY'S VALUES

Creative commons – WWW.ENGAGE-META.COM

## 4. YOUR SITUATION & OBJECTIVES