**Leitfaden zur Erstellung eines Tranzparenz Interfaces**

Das Transparenz Interface stellt textbasiert globale Informationen über ein KI-System zur Verfügung. Es bietet anhand von vier Kategorien, die auf drei verschiedenen Stufen des Informationsumfangs angeboten werden, verständliche Informationen für (potentielle) Endnutzer:innen oder Operator:innen.



Abbildung 1: Übersicht der Kategorien des Transparenz-Interfaces: Ziele, Daten, Art des Systems und Kontrolle, gestaffelt nach Informationsumfang (Stufe 1 bis 3).

**4 Kategorien**

* **Ziel** **des KI- Systems**: Wofür und ich welchem Kontext wird das KI-System eingesetzt?
* **Art des Systems**: Welches System wird eingesetzt und wie funktioniert es? (z.B. regelbasierte Algorithmen, Deep Learning, etc.)
* **Daten**: Mit welchen Trainings- und Betriebsdaten arbeitet das System?
* **Kontrolle**: Welche menschlichen Kontrollmöglichkeiten hat das KI-System und welche Risiken gehen beim Einsatz und dem Grad der Automatisierung einher?

**3 Stufen**

Es wird empfohlen, Nutzer:innen Informationen in unterschiedlichem Umfang anzubieten. Somit kann jede Person je nach Bedarf zu den unterschiedlichen Kategorien mehr oder weniger Informationen anfordern. Um initial einen Informationsüberschuss zu vermeiden, bietet es sich an, Stufe 2 und 3 zunächst in aufklappbare Bereiche einer Website einzubetten und mit „mehr Informationen“ zu betiteln. Wichtig zu beachten ist, dass sich für die verschiedenen Stufen nicht die Komplexität der Sprache ändern sollte. Lediglich der Umfang an Informationen wird erweitert.

1. **Minimaler Informationsumfang**

* In ein bis zwei Sätzen werden die wichtigsten Informationen über die Kategorie in einfacher Sprache beschrieben. Leser:innen sollen die Leitfrage der jeweiligen Kategorie anhand dieses Satzes beantworten können, bekommen aber keine weiteren Hintergründe oder Erklärungen.

1. **Detaillierterer Informationsumfang**

* In drei bis vier Sätzen werden die Informationen der Kategorie zusammengefasst. Der Sprachstil verändert sich im Vergleich zu Stufe 1 nicht. Es werden nur mehr Informationen zu der jeweiligen Kategorie gegeben.

1. **Umfassendste Informationstiefe**

* In Stufe drei werden die meisten Informationen zu einer Kategorie gegeben. Diese werden je nach Kategorie in fünf bis sechs Sätzen abgebildet.

Im Folgenden werden die Kategorien mit drei Stufen des Informationsumfangs beispielhaft an einem KI-basierten Job-Matching-System ausgearbeitet und beschrieben worauf zu achten ist:

Beispiel: KI-basiertes Job-Matching-System

**Ziel des KI Systems**

1. **Stufe 1 (minimal):**

* In einem Satz wird das Ziel des Einsatzes des KI-System benannt und kurz angeschnitten, in welchem Kontext wer von den Ergebnissen betroffen ist.
* „*Das KI-System verfolgt das Ziel, die bestmögliche Lösung für die richtige Übereinstimmung zwischen Bewerber:innen und Stellenanbieter:innen zu bieten, basierend auf den jeweiligen individuellen Bedürfnissen und Qualifikationen.“*

1. **Stufe 2 (detailliert):**

* In der zweiten Stufe wird das Ziel in drei Sätzen erläutert. Hier werden noch weitere Hintergrund-Informationen und untergeordnete Ziele mit aufgeführt.
* *„Das System analysiert detailliert die Präferenzen der Nutzer:innen und verwendet fortschrittliche Algorithmen, um maßgeschneiderte, zielgerichtete Vorschläge zu generieren. Diese Empfehlungen sind optimal auf die spezifischen beruflichen Ziele und Qualifikationen der Nutzer:innen abgestimmt, sodass eine passgenaue und zufriedenstellende Vermittlung gewährleistet ist. Dadurch wird nicht nur die Effizienz erhöht, sondern auch die Wahrscheinlichkeit einer nachhaltigen beruflichen Bindung verbessert.“*

1. **Stufe 3 (umfassend):**

* In der dritten Stufe wird das Ziel des Systems in mit allen zur Verfügung stehenden Kontextfaktoren eingeordnet. Darüber hinaus werden weitere untergeordnete Ziele oder Personengruppen genannt, die potentiell von Entscheidungen der KI betroffen sind. Außerdem werden potentielle Folgen des Einsatzes dieses KI-Systems für ein Unternehmen beschrieben. In dieser Kategorie werden dafür sechs Sätze verwendet.
* *„Das KI-System analysiert durch diese Algorithmen und Methoden des maschinellen Lernens eine Vielzahl an Faktoren, was es ermöglicht, personalisierte Vorschläge zu erstellen, die eine starke Passung sicherstellt. So unterstützt das System effektiv bei der Entscheidungsfindung und ermöglicht nachhaltige berufliche Vermittlungen, die für beide Seiten zufriedenstellend sind. Es soll eine geringere Mitarbeitendenfluktuation und höhere Mitarbeitendenzufriedenheit erreicht werden. Aspekte, die die Arbeitszufriedenheit und Resilienz eines Unternehmens positiv beeinflussen. Die Algorithmen können dabei nicht nur Qualifikationen und Erfahrungen analysieren, sondern auch Soft Skills und Passung zur Organisation einbeziehen, was die langfristige Stabilität und Harmonie in Teams fördert. Dadurch steigert das System die Effektivität der Personalentscheidungen und verbessert nachhaltig die Unternehmenskultur.“*

**Art des Systems**

1. **Stufe 1 (minimal)**:

* In einem Satz wird kurz benannt, dass das KI-System fortschrittliche Algorithmen verwenden. Allerdings werden keine weiteren Informationen zu der genauen Art genannt.
* *„Das KI-System basiert auf fortschrittlichen Algorithmen, die speziell darauf ausgelegt sind, die richtige Übereinstimmung zwischen Bewerber:innen und Stellenanbieter:innen zu unterstützen und zu optimieren.“*

1. **Stufe 2 (detailliert):**

* In drei Sätzen wird die Art des Systems und seine Funktionsweise genauer erläutert.
* *„Es werden deterministische Algorithmen und maschinelles Lernen kombiniert, um präzise und verlässliche Matches zwischen Bewerber:innen und Stellenanbieter:innen zu ermöglichen. Das System passt sich flexibel an individuelle Präferenzen an und gewährleistet eine hohe Genauigkeit bei der Übereinstimmung. Zudem berücksichtigt das System historische Daten und aktuelle Trends, um optimale Ergebnisse zu liefern und gleichzeitig zukünftige Anforderungen zu antizipieren.“*

1. **Stufe 3 (umfassend)**:

* In sechs Sätzen werden neben der Benennung der Systemfunktionen umfangreichere Informationen über die Funktionsweise zur Verfügung gestellt. Darüber hinaus werden Beispiele genannt.

„*Das KI-System erkennt Muster in den Bewerbungen und Stellenausschreibungen und führt passende Profile zueinander. Vor allem standardisierte Eingaben, wie Werdegänge, lassen sich sehr gut einander zuordnen. KI-Bausteine werden nur dann eingesetzt, wenn sie zu individuell besseren Matches führen, z. B. indem sie Ähnlichkeiten in frei formulierten Antworten erkennen. Zusätzlich ermöglicht das System den Bewerbenden, ihr Profil mithilfe eines integrierten Large Language Models (LLM) zu optimieren, indem sie relevante Kompetenzen genauer beschreiben und bestehende Inhalte zusammenführen können. Transparenz steht im Mittelpunkt des Matching-Prozesses: Nutzer:innen erhalten detaillierte Einblicke, warum ein Match vorgeschlagen wurde, und können selbst entscheiden, ob sie dem Vorschlag zustimmen möchten. Erst nach Zustimmung beider Seiten wird das Match finalisiert. Alle Algorithmen und KI-Bausteine sind sorgfältig ausgewählt, um Fairness, Individualität und Diskriminierungsfreiheit zu maximieren und gleichzeitig effiziente und zielgerichtete Matches zu gewährleisten.“*

**Daten**

1. **Stufe 1 (minimal):**

* Es wird in einem Satz beschrieben, woher die im System verwendeten Daten stammen und wofür sie genutzt werden.
* *„Das System verwendet Daten, die aus von Nutzer:innen bereitgestellten Informationen stammen und dienen ausschließlich der Verbesserung der richtigen Übereinstimmung zwischen Bewerber:innen und Stellenanbieter:innen.“*

1. **Stufe 2 (detailliert):**

* In vier Sätzen werden weitere Hintergrundinformationen zu der Herkunft der Daten sowie den Datenschutzrichtlinien gegeben. Außerdem wird darauf eingegangen, wie viel Entscheidungsspielraum Nutzer:innen haben.
* *„Alle Daten werden unter strenger Einhaltung der Datenschutzrichtlinien verarbeitet und sicher gespeichert. Das System zieht nur die Daten heran, die für die Erreichung eines optimalen Matches notwendig sind. Die Nutzer:innen bestimmen dabei selbst, welche Daten in das System eingegeben werden sollen. Für die Erstellung der Matches werden wissenschaftlich fundierte Fragebögen verwendet, welche relevante Kriterien für das Job-Matching enthalten.“*

1. **Stufe 3 (umfassend):**

* In fünf Sätzen wird die verwendete Datenstruktur umfangreicher dargelegt. Darüber hinaus fließen konkrete Beispiele in die Erläuterungen mit ein.
* *„Die Daten werden sicher gespeichert und verarbeitet, um den höchsten Sicherheitsstandards gerecht zu werden. Dabei umfasst das System Informationen der Bewerber:innen, wie beruflicher Werdegang, Ausbildung, Interessen und Kenntnisse, sowie Daten der Stellenanbieter:innen, darunter das Anforderungsprofil und die Rahmenbedingungen. Mithilfe fortschrittlicher Analysetools werden strukturierte und unstrukturierte Daten miteinander verknüpft und analysiert, um präzise und personalisierte Empfehlungen zu generieren. Datenschutz steht dabei an oberster Stelle, sodass eine bestmögliche Übereinstimmung gewährleistet wird, ohne die Sicherheit der Informationen zu gefährden. Zusätzlich bietet das System eine transparente Darstellung der Analyseprozesse, sodass Bewerber:innen und Stellenanbieter:innen nachvollziehen können, wie die Ergebnisse zustande kommen. Regelmäßige Aktualisierungen der Algorithmen gewährleisten zudem, dass das System stets auf dem neuesten Stand der Technik bleibt und optimal arbeitet.“*

**Kontrolle**

1. **Stufe 1 (minimal)**:

* In einem Satz wird genannt, dass es eine menschliche Aufsicht gibt und welche Ziele diese verfolgt.
* *„Eine menschliche Aufsicht stellt sicher, dass das System für die Verbesserung der richtigen Übereinstimmung zwischen Bewerber:innen und Stellenanbieter:innen korrekt funktioniert und faire Entscheidungen trifft.“*

1. **Stufe 2 (detailliert):**

* In dieser Stufewird in vier Sätzen genauer dargelegt, wie die menschlichen Kontrollprozesse ablaufen und welche Möglichkeiten eine menschliche Aufsicht hat, um Risiken rechtzeitig zu erkennen und zu beheben.
* *„Regelmäßige Audits und Testverfahren sorgen dafür, dass das System keine unvorhergesehenen Ergebnisse liefert. Die Überwachung erfolgt durch eine Kombination aus menschlicher Aufsicht und automatisierten Kontrollmechanismen, die Unregelmäßigkeiten frühzeitig erkennen. Dabei werden Datenprotokolle kontinuierlich analysiert, um Muster und potenzielle Risiken zu identifizieren. Zusätzlich stellen flexible Anpassungsmöglichkeiten sicher, dass das System auf veränderte Anforderungen oder neue Szenarien schnell reagiert.“*

1. **Stufe 3 (umfassendste Stufe):**

* In fünf Sätzen werden neben den in Stufe 2 genannten Informationen Beispiele aufgeführt.
* „*Die menschliche Aufsicht umfasst Überprüfungen, z. B. von Testfällen eines Matches, und wird durch automatisierte Überwachungsmechanismen sowie die Überprüfung von Grenzwerten ergänzt. Die Systemarchitektur integriert ein Audit-System, das sämtliche Entscheidungen nachvollziehbar macht und alle relevanten Daten für die Kontrolle bereitstellt. Bei Abweichungen oder Anomalien, wie Biases in der Matching-Entscheidung, können Expert:innen direkt eingreifen und Anpassungen vornehmen. Dadurch bleibt das System transparent und gewährleistet, dass alle Entscheidungen fair und nachvollziehbar sind. Zusätzlich wird die Systemleistung regelmäßig evaluiert, um sicherzustellen, dass aktuelle Standards und regulatorische Anforderungen erfüllt werden und um das System langfristig zu stärken“*