



René Riedl

Vertrauen in der digitalen Welt der Zukunft

In welchen Bereichen der digitalen Zukunft müssen wir worauf vertrauen können?

Wenn wir uns mit Vertrauen in der digitalen Welt befassen, ist zuerst einmal die Feststellung wichtig, dass „die digitale Welt“ etwas sehr Heterogenes ist. Beispiele sind: Haben wir Vertrauen beim Online-Shopping? Wenn ja, in wen genau. Beispielsweise in die Plattformen Amazon oder eBay oder den konkreten Verkäufer, entweder als Privater oder als Unternehmen. Vertrauen wir unseren Freunden auf Facebook? Haben wir Vertrauen in die Sicherheit unserer Daten, wenn diese in der Cloud gespeichert werden, also auf Servern von Unternehmen, auf die wir über Internet zugreifen? Wie vertrauenswürdig ist ein potenzieller Partner, dessen Foto wir auf einer Dating-Plattform sehen? Vertrauen wir neuen Technologien wie der Blockchain und darauf aufbauenden Anwendungen wie z. B. der Kryptowährung Bitcoin? Vertrauen wir Künstlicher Intelligenz? Und welcher Manifestation von Künstlicher Intelligenz? Sprachassistenten wie Amazons Alexa oder Apples Siri? Selbstfahrenden Autos? Einem Computerprogramm, das auf der Basis der Eingabe von Symptomen eine Diagnose stellt? Oder vertraut ein Radiologe auf eine computerbasierte Befundung von Kernspintomographiebefunden?

Diese und viele weitere Vertrauensfragen stellen sich in einer digitalen und maschinenbasierten Welt. Und an diesen Beispielen sehen wir, wie unterschiedlich Vertrauenssituationen im Internet sein können. Um solche Situationen besser verstehen zu können, hat man in der Wissenschaft eine Systematik von Vertrauenssituationen entwickelt, die zumindest drei Szenarien unterscheidet. Anhand dieser Systematik kann man gut nachvollziehen, welche Voraussetzungen gegeben sein müssen, um Vertrauen aufzubauen oder zu pflegen.

Szenario 1: Vertrauen in andere Menschen (z. B. bei computerbasierter Interaktion wie Käufer-Verkäufer auf eBay oder Online-Dating Plattformen). Die drei Faktoren, die hier die Vertrauenswürdigkeit eines Interaktionspartners stark beeinflussen, sind:

- ❖ **Kompetenz** (Welche Fähigkeiten und Fertigkeiten hat die andere Person, die für unsere Interaktion und Beziehung notwendig sind?)
- ❖ **Wohlwollen** (Wie sehr interessiert sich und berücksichtigt die andere Person meine eigenen Interessen?)
- ❖ **Integrität** (Wie sehr hält sich die andere Person an bestehende Regeln, Werte, Normen und Gesetze?)

Szenario 2: Vertrauen in IT-Systeme (z. B. Software am Arbeitsplatz, Cloud-Anwendungen). Die drei Faktoren, die hier die Vertrauenswürdigkeit eines IT-Systems stark beeinflussen, sind:

- ❖ **Funktionalität** (In welcher Weise und in welchem Ausmaß bietet das IT-System Features an, die mich bei der Erfüllung meiner Aufgaben unterstützen?)
- ❖ **Support** (Wie sehr unterstützt mich das IT-System bei Problemen? Geht es auf meine Anforderungen und meinen Kenntnisstand ein?)
- ❖ **Zuverlässigkeit** und Prognostizierbarkeit des Systemverhaltens (Ist das System zuverlässig, stabil und sein Verhalten vorhersagbar?)

Szenario 3: Vertrauen in autonome Systeme und künstliche Intelligenz (z. B. Sprachassistenten, autonom fahrende Autos). Die drei Faktoren, die hier die Vertrauenswürdigkeit eines IT-Systems stark beeinflussen, sind:

- ❖ **Performance** (Wie sehr unterstützt mich die Technologie bei der Erreichung meiner Ziele, beispielsweise, um automatisiert bestimmte Prozesse zu erledigen?)
- ❖ **Zweck** (Wird die Technologie dem Verwendungszweck entsprechend angewendet?)
Beispiel Sprachassistent: Es wird der Befehl zum Abspielen bestimmter Musikstücke gegeben versus das Mithören und Aufzeichnen von Gesprächen (Stichwort: Verletzung der Privatsphäre).
- ❖ **Nachvollziehbarkeit** der Entscheidungen und des Technologieverhaltens (z. B. Fahrverhalten autonomer Fahrzeuge, automatischer erstellte Befunde in der Medizin auf der Basis von Künstlicher Intelligenz).

Welche Chancen für die Zukunft sehen Sie für jene, die sich durch Vertrauenswürdigkeit im Bereich Digitalisierung auszeichnen?

Vorab ist hierzu zu sagen, dass der Mensch eine natürliche Tendenz in Richtung Vertrauen hat, ich spreche hier vom sogenannten Urvertrauen. Dies wird in der Fachliteratur mitunter evolutionär erklärt, da der menschliche Nachwuchs nach der Geburt nicht selbständig lebensfähig ist (im Vergleich zu bestimmten Tierarten). Daraus folgt, dass ein Baby seiner sozialen Umgebung – insbesondere natürlich der Mutter – zugeneigt ist, und auch umgekehrt die Mutter dem Baby. Es sind hier mehrere Hormone im Spiel, insbesondere Oxytocin, das auch als Bindungs- und Vertrauenshormon bezeichnet wird. Zudem weiß man heute, dass ein gewisser Teil des menschlichen Vertrauensverhaltens durch das Zusammenspiel von Umwelt und genetischer Prädisposition erklärt werden kann, also ein epigenetisches Phänomen ist. Dennoch ist Vertrauen situationsspezifisch beeinflussbar und es geht mit einer Vielzahl positiver Wirkungen einher: Ökonomen berichten, dass in Ländern, in denen sich Menschen mehr Vertrauen, die Armut geringer

und die wirtschaftliche Leistungsfähigkeit eines Landes besser ist, weil mehr wirtschaftliche Transaktionen stattfinden. Psychologen und Soziologen berichten, dass höheres Vertrauen mit offenerem Verhalten in Interaktionssituationen und Beziehungen einhergeht, dass mehr Informationen geteilt werden, dass man überhaupt mehr aufeinander zu- und eingeht, und dass das Wohlbefinden besser ist, als in Situationen von Misstrauen.

Daraus folgt, dass Unternehmen danach streben sollten, Vertrauen zu signalisieren und vertrauenswürdig zu sein. Je nach Situation kann ein solches Vertrauen über die zuvor beschriebenen Faktoren induziert werden (vgl. Szenario 1, 2, 3). Im Bereich der Künstlichen Intelligenz haben sich zudem spezifische Vertrauensfaktoren herauskristallisiert:

- ❖ Nutzen für das **Wohlergehen der Gesellschaft** als Gesamtes (und nicht nur einzelner Akteure und Unternehmen),
- ❖ **Vermeidung** von Situationen, in denen für Menschen **Schaden** entstehen kann,
- ❖ **Autonomie** des Menschen, woraus folgt, dass der Mensch immer die Letztentscheidung hat (man denke hier an medizinische Diagnosen und Therapieempfehlungen durch Künstliche Intelligenz oder an autonomes Fahren),
- ❖ **Gerechtigkeit** im Sinne von Verteilungsgerechtigkeit des Nutzens durch Künstliche Intelligenz,
- ❖ **Nachvollziehbarkeit** von Entscheidungen durch Künstliche Intelligenz (was kein einfaches Unterfangen ist, da Entscheidungen auf der Basis des maschinellen Lernens und Deep-Learning-Verfahren das Wesen einer Black Box haben).

Abschließend kann gesagt werden, dass Vertrauen in der digitalen Welt ein zentrales Thema ist, das in Wissenschaft und Praxis in Zukunft noch weiter an Relevanz gewinnen wird. Das Funktionieren digitaler Wirtschaft und Gesellschaft basiert entscheidend auf Vertrauen.