

State of the Art

Die Überwachung von Mitarbeitern ist tatsächlich gar nicht so selten wie man vielleicht vermuten möchte. Zwar gibt es hierzu in Österreich einigermaßen gute Gesetze, so ist die Überwachung nur mit Zustimmung des Betriebsrates bzw. der betroffenen Mitarbeiter erlaubt. Außerdem müssen die Daten nach strengen Richtlinien gespeichert werden. Allerdings gibt es einige Länder, in denen diese Regelungen weitaus weniger strikt sind.

Viele der gängigen Kollaborations-Tools, die in Unternehmen eingesetzt werden, wie etwa Microsoft Teams, bieten auch schon dementsprechende Tools an. Mit MS Teams lassen sich beispielsweise Chat-Analysen erstellen, Meetings-Transkripte erstellen und auswerten oder sogar das gesamte „digitale“ Verhalten der Mitarbeiter. Dadurch versucht Teams herauszufinden, welche Mitarbeiter eine potenzielle Bedrohung für das Unternehmen in irgendeiner Form darstellen könnten.

Creative

Bei der Ideenfindung standen wir zu Beginn vor der Challenge, dass Policy Thinking in der VO noch nicht behandelt wurde. Wir hatten daher noch keinen konkreten Plan, wie wir das ordentlich mit Critical Thinking verbinden können unsere Überlegungen fokussierten sich daher hauptsächlich auf das Critical Thinking und wie wir dies mit gewissen Richtlinien verbinden können. Zum Thema Critical Thinking wollten wir etwas, entwickeln, dass einerseits einen Nutzen für eine Partei hat, andererseits aber durchaus kritisch hinterfragt werden kann und sollte. Durch den aktuellen Hype rundum KI kamen wir sehr schnell auf die Idee etwas durch KI analysieren zu lassen und da Überwachung prinzipiell immer ein kritisches Thema ist fanden wir es spannend eine überwachte Chat Anwendung für Mitarbeiter zu entwickeln.

Normstorming

	MS Teams Compliance Tools	Slack Moderation Bots	Discord Moderation Bots
Integration	Integriert	Externes Tool / Erweiterung	Externes Tool / Erweiterung
Anpassbarkeit	Stark	Stark	Stark
Fokus	Compliance, Corporations	Compliance, Agile Teams	Privatpersonen, Gaming
Gemeinsame Normen	Schutz der Nutzer, Förderung der Produktivität, Data Loss Prävention	Schutz der Nutzer, Förderung der Produktivität	Schutz der Nutzer, Verhinderung von Hate Speech

Alle drei Alternativen haben gemeinsame Werte in gewissen Aspekten, sie versuchen Diskriminierung über Chats zu verringern. Der Ansatz wie das geschehen soll ist

teilweise unterschiedlich. So hat MS Teams viele integrierte Tools, was durchaus Sinn macht, da man dadurch wenig Mehraufwand hat. Slack und Discord bieten allerdings viel mehr Erweiterungen und somit auch Eigenentwicklungsmöglichkeit, was Sinn macht, wenn es um sehr sensible Daten geht.

Fame vs Shame

	Bewertung	Grund
MS Teams Compliance Tools	Gut	Gut integrierbar, langfristiger Support und ständige Updates, starker Fokus auf Datenschutz (!!)
Slack Moderations Bots	Gut	Stark personalisierbar, vielseitig, benutzerfreundlich für kleinere und mittlere Teams
Discord Moderation Bots	Schlecht	Wenig Fokus auf Datenschutz bei Dritt-Tools, Ursprünglich für Gaming-Sektor entwickelt

Beim Mind-Mapping haben wir ein paar verschiedene Ansätze betrachtet, wie wir dieses Konzept umsetzen könnten.

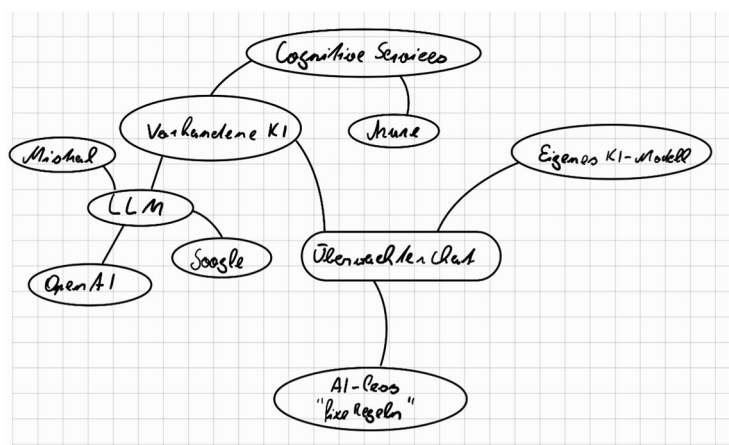


Abbildung 1 Mind-Mapping

Die Entscheidung ist schlussendlich klar auf eine Umsetzung mittels dem LLM von OpenAI (ChatGPT) gefallen. Alle anderen Varianten waren entweder zu aufwendig (eigenes KI-Modell) oder hatten weniger Aussichten auf Erfolg (AI-less Approach bzw. Cognitive Services). Von den LLMs hatte OpenAI die am besten dokumentierte API, daher entschieden wir uns dann für diesen Weg.

Im Zuge der Lotus blossom Übung haben wir uns noch genauer überlegt, was Teil unseres Projekts sein sollte. Genauere Beschreibung der einzelnen Punkte siehe „Project“.

Definition von Chat - Richtlinien	Eigenes Chat-Tool für MA	Echtheit - Analyse der Chats von den MAs
Analyse der Vielsamkeit	Chat-Analyse Software für Analyse von firmeninternen Chats	Einsparung „verdächtig“ Chats in HR Postfach
Vergleich zw. menschlicher u. maschineller Analyse	Vergleich zweier AI-Systeme	Annotiere zum Verhalten des Systems.

Abbildung 2 Lotus blossom

Project

Unser Projekt war die Umsetzung einer webbasierten Chat Anwendung für Mitarbeiter eines Unternehmens. Die Besonderheit dabei ist, dass die Chats der Mitarbeiter von einem KI-Modell analysiert werden, um herauszufinden, ob es zwischen den Mitarbeitern Konflikte gibt. bzw. sensible Unternehmensdaten ausgetauscht werden. Diese kleine Software soll als Beispiel für die technische und vor allem äußerst fragwürdige Implementierung von Richtlinien dienen.

Für die Analyse wurde die API von OpenAI verwendet. Das AI-Modell bekommt per Prompts Richtlinien für die Chats und soll diese basierend auf diesen analysieren. Zuerst wollten wir die Analyse des AI-Modells sehen, ohne dass wir es zu sehr einschränkten. Daher hat es nur folgende Richtlinie erhalten:

“The output structure must be a valid JSON object with a structure like: { \"ConflictPotential\": double, \"SensitiveLeak\": double, \"CriticalPhrases\": string[] } Your actual task is: You read the chats of employees to find potential foul language, use of swear words, conflict potential of leakage of confidential information. After reading you should analyse and estimate how likely this text is critical in any of the mentioned areas. Highlight any sections that support your reasoning.“

Für folgende Konversation:

Alex: Jamie, why did you take credit for the Johnson project? We agreed it was a team effort.

Jamie: I handled the final report and client communications. That's what matters most.

Alex: The data analysis I did was crucial. You can't just ignore that.

Jamie: The client doesn't see the behind-the-scenes work. They see the final product.

Alex: That's not fair. We agreed to share the credit equally.

Jamie: Maybe you're overreacting. I stand by what I did.

Alex: This isn't right. I won't let it slide this time.

Erhält man damit die folgende Analyse:

```
{
  "ConflictPotential": 0.7,
  "SensitiveLeak": 0.0,
  "CriticalPhrases": [
    "credit for the Johnson project",
    "agreed it was a team effort",
    "handled the final report and client communications",
    "client doesn't see the behind-the-scenes work",
    "We agreed to share the credit equally",
    "I won't let it slide this time"
  ]
}
```

Hier kann man sehen, dass das KI-Modell einen möglichen Konflikt zwischen den beiden Mitarbeitern für wahrscheinlich hält. Aus den Chats gehen allerdings keine sensiblen Unternehmensdaten hervor. Gibt man dem Modell genauere Richtlinien für die Analyse der Chats, dann ändert sich auch sein Output.

Nun haben wir spezifiziert, dass wir vor allem auf Schimpfwörter achten und Diskussionen, die sich möglicherweise nach einem Konflikt anhören, aber sachlich geführt sind und nur auf unterschiedliche Meinungen schließen lassen nicht in die Bewertung einfließen lassen wollen:

"The output structure must be a valid JSON object with a structure like: { \"ConflictPotential\": double, \"SensitiveLeak\": double, \"CriticalPhrases\": string[] } Your actual task is: You read the chats of employees to find potential foul language, use of swear words, conflict potential of leakage of confidential information. The conflict potential estimation should mainly focus on the presence of foul word or abusive language in general. If two employees have a different point of view on something but discuss it in appropriate manner, we do not talk about conflict potential. After reading you should analyse and estimate how likely this text is critical in any of the mentioned areas. Highlight any sections that support your reasoning."

```
{
  "ConflictPotential": 0.2,
  "SensitiveLeak": 0.0,
  "CriticalPhrases": [
    "Jamie, why did you take credit for the Johnson project?",
    "Maybe you're overreacting.",
    "I won't let it slide this time."
  ]
}
```

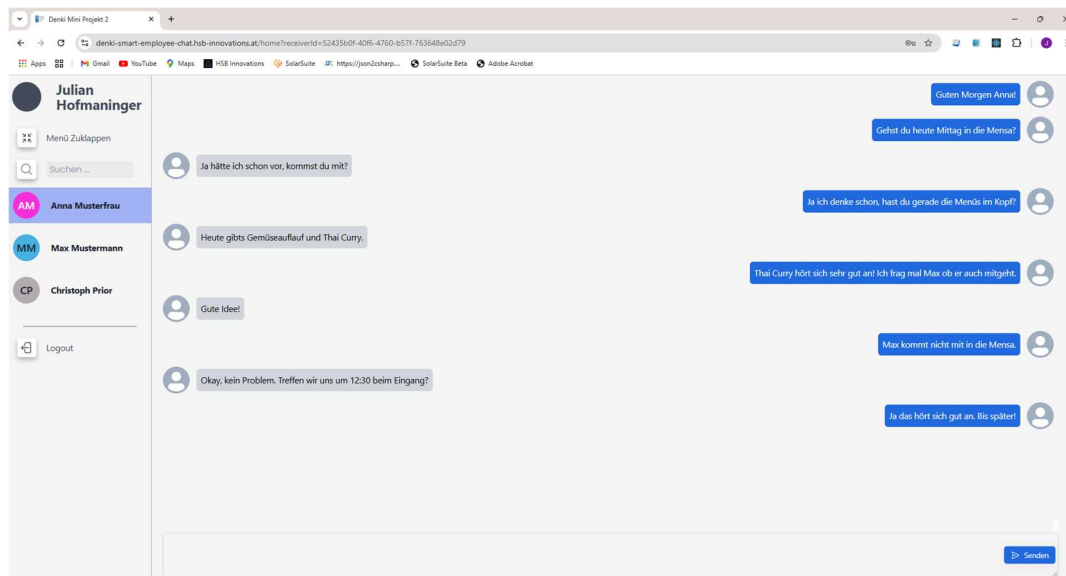


Abbildung 3 Normaler Chat

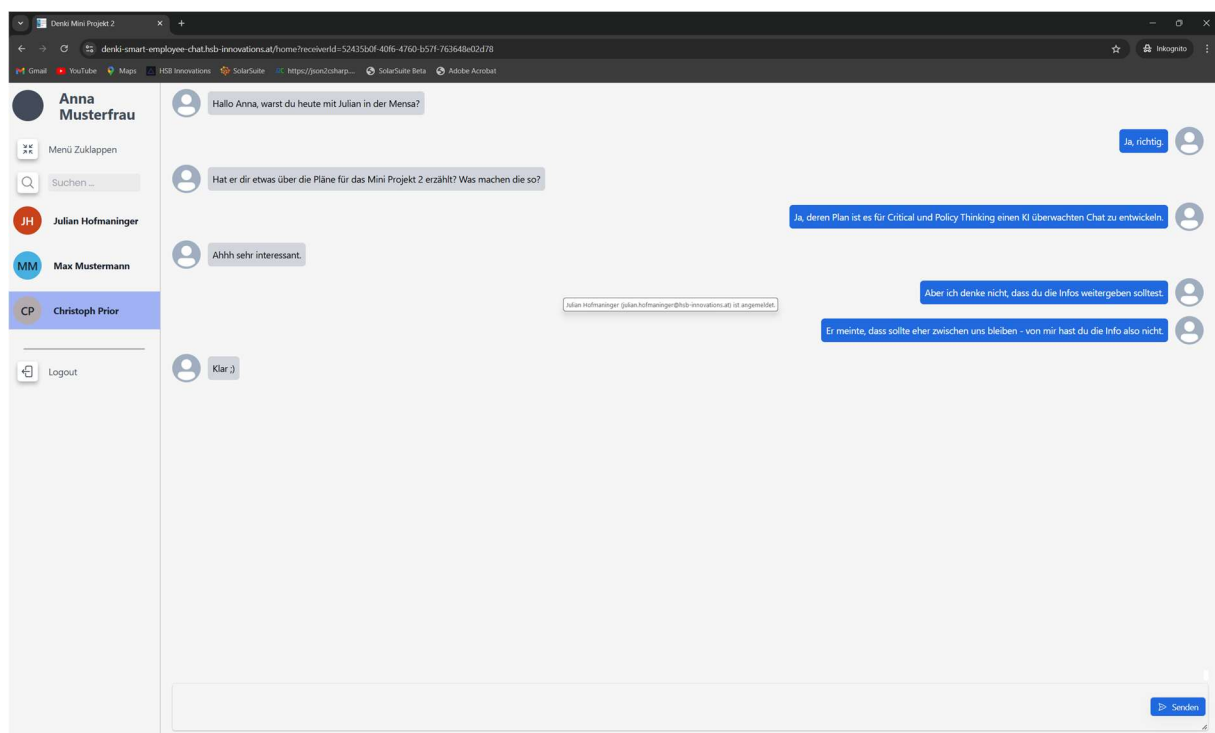


Abbildung 4 Kritischer Chat

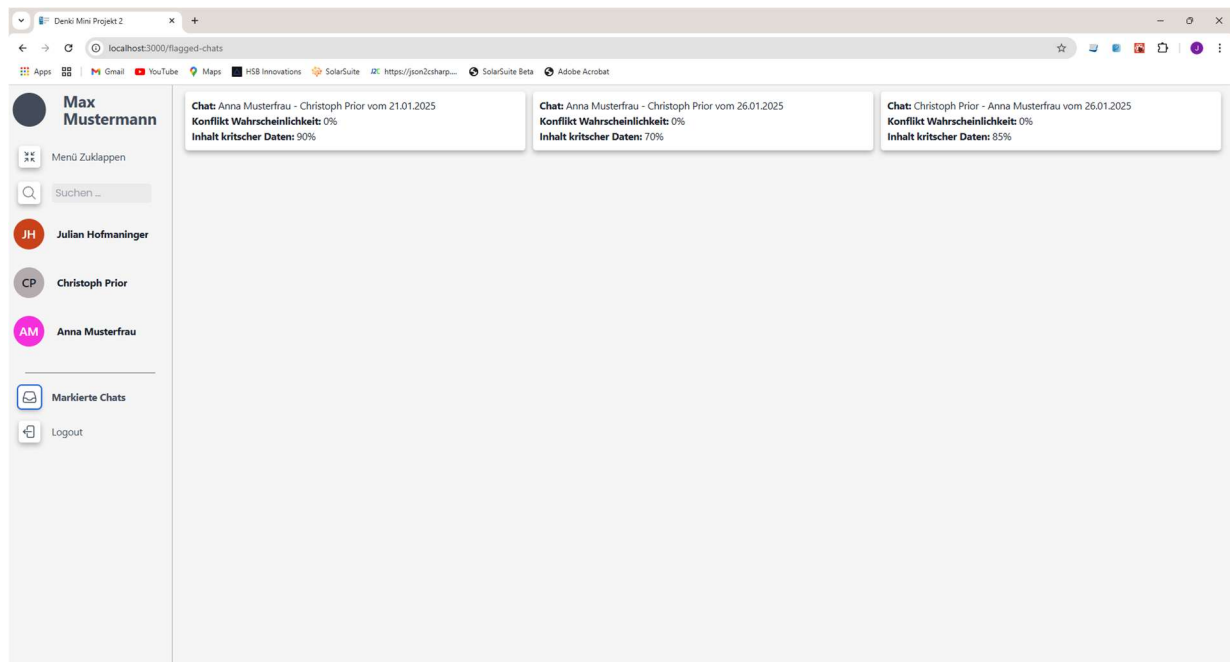


Abbildung 5 Chat Auswertung

Reflexion und Feedback

Feedback durch Dritte

Wir haben drei Personen zu unserem Miniprojekt befragt:

- Stefan, 20, Student
- Max, 35 und Rechtsanwalt
- Daniel, 45 und Unternehmer

Das ist leider nur eine sehr kleine Auswahl an Personen, daher gibt es einige nicht repräsentierte Gruppen wie Frauen, Arbeitnehmer, kulturelle Minderheiten, etc. Dennoch war das Feedback interessant und wertvoll.

„Wäre ich in der Position und mein Unternehmen oder mein Arbeitgeber gezwungen solche Maßnahmen zusetzen würde ich dies erstens nicht befürworten und zweites stark hinterfragen an welchen Stellen hier sonst noch etwas schief läuft. Denn all die Benefits, welche ein solches Tool bringen könnte, sollte man auch durch gutes Leadership erreichen“ - Stefan

„Für Unternehmen bietet eine solche Technologie natürlich sehr interessante Daten. Einerseits kann das Arbeitsklima im Unternehmen überwacht werden, was potenziell zur Erhöhung der Produktivität führt. Andererseits kann man verhindern, dass sensible Daten ausgetauscht werden. Andererseits ist es ethisch stark zu hinterfragen, ob eine solche Überwachung der Mitarbeiter sinnvoll ist und ob diese nicht ohnehin umgangen werden würde. Außerdem ist die rechtliche Grundlage für ein solches Vorhaben etwas unklar.“ – Max

„Ich gehe davon aus, dass meine Mitarbeiter die Ressourcen ohnehin nur zu unternehmensdienlichen Zwecken nutzen und private Konversationen nicht über Firmendienst abgewickelt werden. Implementiert man eine solche Lösung also ordentlich und datenschutzrechtlich korrekt sehe ich nicht, wo das ein Problem darstellen sollte“ - Daniel

Wie wurde Ihr Verständnis der gewählten Denkweisen durch diese Übungsarbeit verändert?

Ich bin nicht der Meinung, dass diese Gruppenarbeit das Verständnis der Denkweisen noch großartig verändert. Entweder man versteht die Grundkonzepte bereits und ist in der Lage diese in einem Projekt anzuwenden oder man muss sich noch etwas einarbeiten.

Inwiefern kann ein nachhaltiges Verständnis der gewählten Denkweisen Ihnen im Studium oder danach im Beruf helfen?

Kritisches Denken ist in vielen Bereichen sehr hilfreich, da es ein erlaubter Zusammenhang zu hinterfragen und dadurch besser zu verstehen. Policy Thinking hingegen hat meiner Meinung nach einen eher beschränkteren Anwendungsbereich. Es könnte aber durchaus hilfreich sein, wenn man vor der Challenge steht Richtlinien für etwas festzulegen. Beispielsweise die Richtlinien für die Verwendung von genAI in einem IT-Unternehmen o.Ä.

Welche Teile dieser Arbeit fanden Sie besonders schwer, welche zu einfach?

Bei dieser Aufgabe war es besonders schwierig ein Projekt zu finden, welches die beiden Denkweisen Critical und Policy richtig verbindet. Meiner Meinung nach ist vor allem Policy Thinking ein sehr theoretisches Thema, was es erschwert eine praktische Anwendung dafür zu finden.

Sind Sie mit Ihrer Arbeit zufrieden?

Ja, ich finde unser Projekt ist insgesamt gelungen. Es war einerseits lustig umzusetzen und andererseits verbindet es wichtige Aspekte der beiden Denkweisen.

Zusätzliche Quellen

<https://www.heise.de/hintergrund/Die-digitale-Buechse-der-Pandora-Wie-Konzerne-ihre-Mitarbeiter-ueberwachen-10202107.html>