

## 7. Symposium Steuern & Bilanzen

# Kann Generative AI das menschliche Urteil des Abschlussprüfers nachbilden und unterstützen?

The logo for WU (Wirtschaftsuniversität Wien) features the letters 'WU' in a large, bold, black serif font.

**WIRTSCHAFTS  
UNIVERSITÄT  
WIEN VIENNA  
UNIVERSITY OF  
ECONOMICS  
AND BUSINESS**

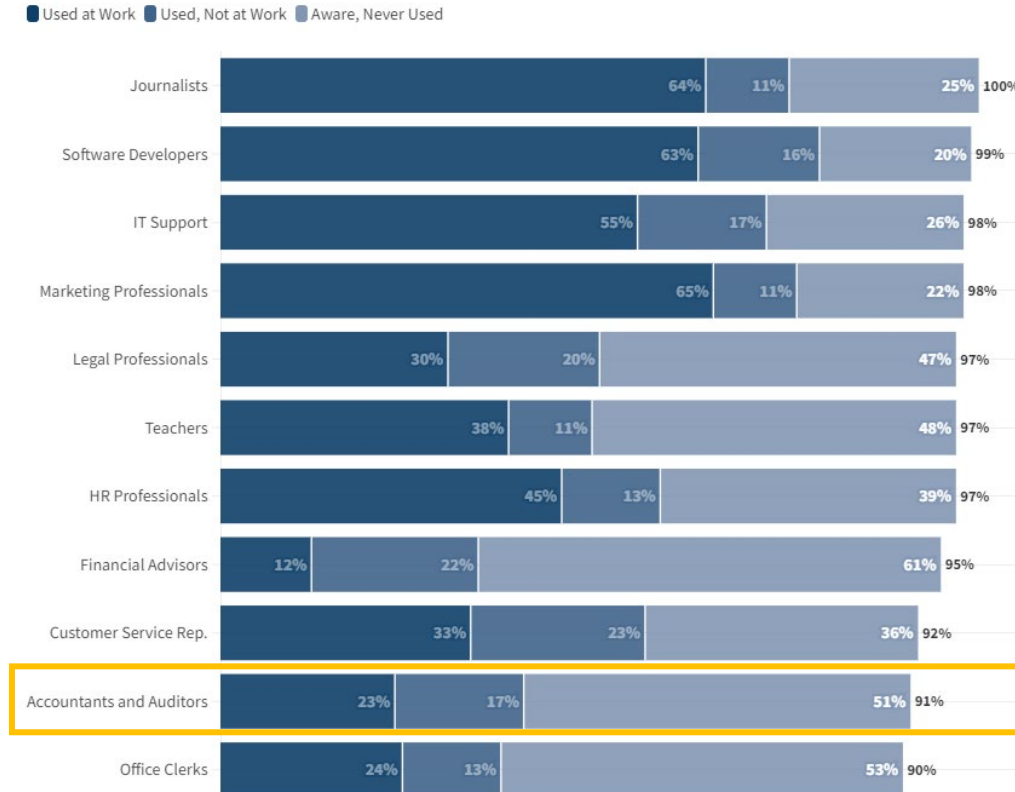
### **Dr. Markus Isack, MSc**

Wirtschaftsuniversität Wien,  
Abteilung für Unternehmensrechnung und  
Revision

Johannes Kepler Universität Linz,  
LIT Open Innovation Center

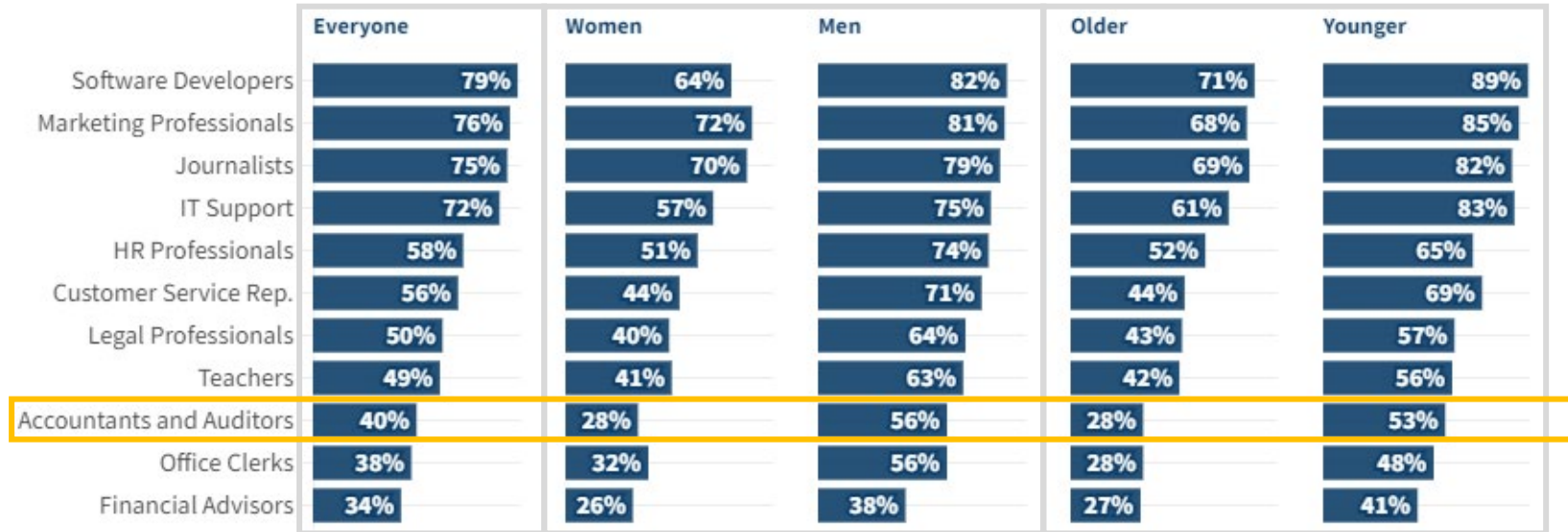
# Nutzung von ChatGPT in der Praxis

(Humlum & Vestergaard, 2024)



# Nutzung von ChatGPT in der Praxis

(Humlum & Vestergaard, 2024)



# Aufgaben im Rechnungswesen und Wirtschaftsprüfung zusammengeführt mit ChatGPT Fähigkeiten

(Humlum & Vestergaard, 2024; Eloundou et al., 2023)

- Ansicht, dass Generative AI bei 38% der Arbeitsaufgaben rund 50% der Arbeitszeit eines *durchschnittlichen* Mitarbeiters reduzieren kann. Im *eigenen* Arbeitsumfeld wird dies für 30% der Arbeitsaufgaben angesehen.
- Beispiele
  - Detaillierte Berichte über Prüfungsergebnisse erstellen
  - Überwachung der Prüfung von Objekten und Festlegung des erforderlichen Prüfungsumfangs
  - Prüfung und Bewertung von Finanz- und Informationssystemen, mit Empfehlungen zur Sicherstellung der Systemzuverlässigkeit und Datenintegrität
  - Gespräche mit Unternehmensvertretern über finanzielle und regulatorische Angelegenheiten führen
  - Erstellen, Prüfen oder Analysieren von Buchhaltungssunterlagen, Jahresabschlüssen oder anderen Finanzberichten zur Bewertung der Genauigkeit, Vollständigkeit und Einhaltung von Berichts- und Verfahrensstandards
  - Berechnung von Steuern und Erstellung von Steuererklärungen unter Einhaltung der Zahlungs-, Melde- oder anderer Steueranforderungen

- Zahlreiche aktuelle Forschungsarbeiten haben die Vielseitigkeit generativer KI in verschiedenen Bereichen wie Medizin, Beratung und Recht unterstrichen (Gu et al. 2024) -> CoPilots in diesen Bereichen
- Generative KIs können jetzt Aufgaben erfüllen, die früher als unerreichbar galten (Tiku et al. 2023; Wei et al. 2022)
- Gespaltene Meinungen in der Wirtschaftsprüfung: besser geeignet für nicht-urteilbezogene -professional judgement- Aufgaben (Fotoh und Mugwira 2023) vs. Vision eines Co-Piloten von Gu et al. (2024)

*"The right way to think of the models that we create is a reasoning engine, not a fact database"*  
Sam Altman, CEO OpenAI

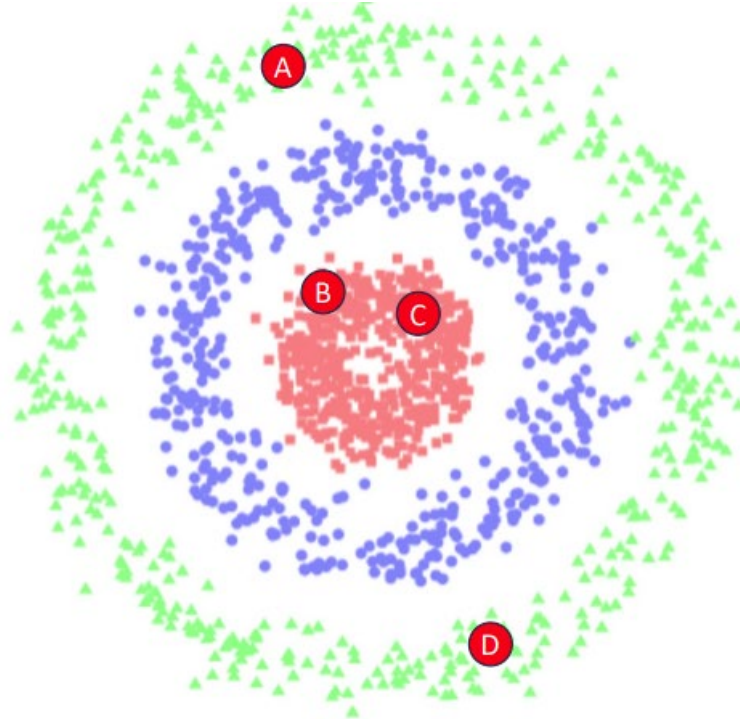
*"I don't know if LLMs can reason in theory, but they seem to "reason" in practice."*  
Ethan Mollick, Co-Director Generative AI Lab at Wharton

- Untersuchung des „Verhaltens“ von GPT-4 in fünf verschiedenen Experimenten, die zuvor mit Wirtschaftsprüfern und Mitgliedern von Prüfungsausschüssen durchgeführt wurden, und Vergleich der Ergebnisse von GPT-4 mit denen der menschlichen Teilnehmern.
  - Angemessenheit der Annahme des Going Concern (Hoos et al. 2019)
  - Fraudererkennung (DeZoort and Harrison, 2018)
  - Beurteilung einer vom Wirtschaftsprüfer vorgeschlagenen Anpassung durch den Prüfungsausschuss (DeZoort et al., 2008)
  - Vertrauen des Wirtschaftsprüfers in die Arbeit der Innenrevision (Malaescu and Sutton, 2015)
  - Wesentlichkeitsanalyse (DeZoort et al., 2006)

- Das Prinzip generativer KIs wie GPT-4 besteht darin, Text zu erzeugen, der als ‚vernünftige Fortsetzung‘ eines gegebenen Textes beschrieben werden kann.
  - Modelle replizieren nicht einfach Texte; sie interpretieren und vergleichen Bedeutungen (basierend auf wahrscheinlichkeitsbasierten Parametern -> Ähnlichkeit) und erzeugen kontextuell passende Antworten.
  - Variablen wie *temperature* und *seed number* führen zu Zufälligkeiten -> menschlichere und nicht wiederholbare Ergebnisse
  - Die Mechanismen, die generativen AIs zugrunde liegen, sind nicht vollständig verstanden (Keine „Programmierung“ im Lernprozess – es sind Berechnungen!)

# Ähnlichkeit – Abhängig vom Kontext

(Nessler, 2024)

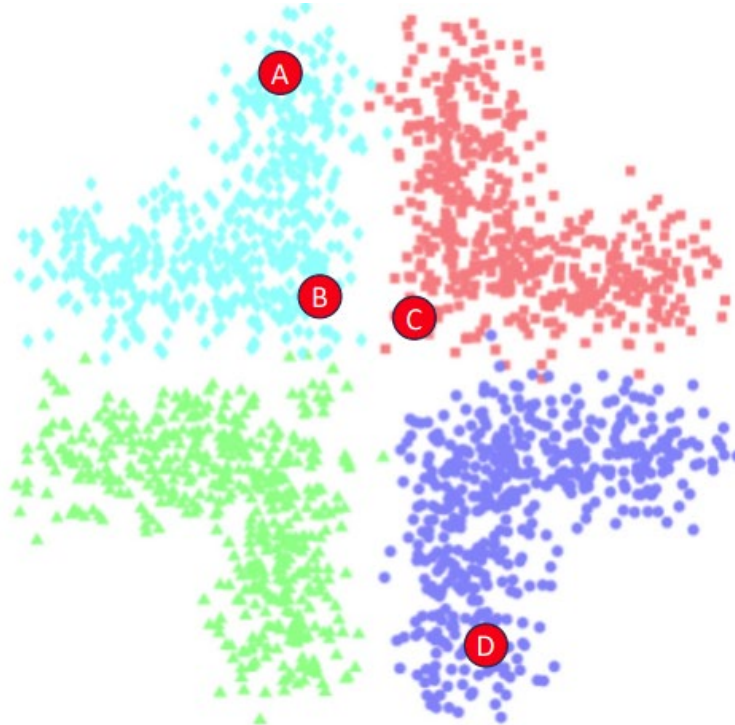


B,C ähnlich  
A,D ähnlich



# Ähnlichkeit – Abhängig vom Kontext

(Nessler, 2024)



A,B ähnlich  
B,C verschieden



- Die Forschung zu Generative AI untersucht die Fähigkeiten und Limits (wie z.B. die Red-Team-Forschung von OpenAI)
- Vergleich der Antworten menschlicher Wirtschaftsprüfer und Mitglieder von Prüfungsausschüssen mit denen von GPT-4 (z.B. Eriksen et al. (2023), Dillion et al. (2023), Bubeck et al. (2023), Berg et al. (2023), und Collins et al. (2023))
- Verwendung desselben „Prompts“ für beide Gruppen, wie es in Replikationsstudien üblich ist (Demszky et al. 2023)
  - Zero-Shot-Ansatz (kein Fine-Tuning) zur Bewertung der zugrunde liegenden Fähigkeiten (keine Verzerrung durch Fine-Tuning, wie von Röttger et al. (2024) vorgeschlagen)
  - 10-fache neugenerierung zur Kontrolle potenzieller Zufälligkeiten (z.B. aufgrund der *temperature*)
  - Leakage-Test ergab keine Hinweise darauf, dass die experimentellen Fälle Teil der Trainingsdaten von GPT-4 waren
- Die Auswahl der fünf Studien erfolgte auf Basis a) eines breiten Spektrums an Prüfungsaufgaben und b) methodischer Transparenz (bereitgestellte vollständige experimentelle Aufgaben!)

**You**

Please assume that you are a member of Wareham Manufacturing, Inc.'s (Wareham) audit committee. Wareham has just completed its fiscal year. This case focuses on an unresolved accounting issue that will be discussed at the next audit committee meeting. The meeting will be held prior to publicly releasing annual financial information.

#### Background

Wareham is a publicly traded company that manufactures and distributes cellular telephones to retailers throughout the eastern half of the United States. Wareham's management recognizes that heavy competition in the industry and frequent technological advances make inventory obsolescence and inventory valuation key business and accounting issues.

Wareham has been audited by the same Big 5 accounting firm for the preceding five years.

Wareham's auditor has always issued standard, unmodified audit reports. All quarterly financial information has been reviewed on a timely basis by the same accounting firm. Since you joined the audit committee, the committee has regularly held quarterly meetings with management and the auditor. No significant disagreements between management and the auditor have reached the audit committee during your tenure.

Management has notified the audit committee that it is not inclined to increase its inventory write-down to the auditor's estimate because they consider the adjustment to be immaterial. The auditor is inclined to agree with management and recommends that management should not adjust the inventory write-down.

Based only on the information presented on the preceding case provide answers from the perspective of a member of Wareham Manufacturing, Inc.'s (Wareham) audit committee:  
What do you believe should be the final EPS figure? Provide a number within the range from \$1.07 to \$1.10

What is your opinion with respect to the auditor's proposed adjustment to management's inventory write-down? Provide a number within the scale 0 to 100 (0 = Definitely do not

recognizes that heavy competition in the industry and frequent technological advances make inventory obsolescence and inventory valuation key business and accounting issues. Wareham has been audited by the same Big 5 accounting firm for the preceding five years. Wareham's auditor has always issued standard, unmodified audit reports. All quarterly financial information has been reviewed on a timely basis by the same accounting firm. Since you joined the audit committee, the committee has regularly held quarterly meetings with management and the auditor. No significant disagreements between management and the auditor have reached the audit committee during your tenure.

Management has notified the audit committee that it is not inclined to increase its inventory write-down to the auditor's estimate because they consider the adjustment to be immaterial. The auditor is inclined to agree with management and recommends that management should not adjust the inventory write-down.

Based only on the information presented on the preceding case provide answers from the perspective of a member of Wareham Manufacturing, Inc.'s (Wareham) audit committee:  
What do you believe should be the final EPS figure? Provide a number within the range from \$1.07 to \$1.10

What is your opinion with respect to the auditor's proposed adjustment to management's inventory write-down? Provide a number within the scale 0 to 100 (0 = Definitely do not adjust; 100 = Definitely adjust)

What do you think the typical audit committee member's opinion about the auditor's proposed adjustment would be? Provide a number within the scale 0 to 100 (0 = Definitely recommend to not adjust; 100 = Definitely recommend to adjust)

What do you think the final outcome of the auditor's adjustment would be? Provide a number within the scale 0 to 100 (0 = Definitely will not be adjusted; 100 = Definitely will be adjusted)

What is your rationale for the judgment concerning whether the auditor's proposed adjustment should be recorded?

 Message ChatGPT



To address the scenario provided as a member of Wareham Manufacturing, Inc.'s audit committee, I will analyze the given information and provide a structured response to each of the questions posed.

## 1. Final EPS Figure

- **EPS Range:** \$1.07 to \$1.10
- **Determining Factors:** Auditor's proposal vs. Management's stance, materiality threshold, potential market impact.
- **Suggested EPS:** Given that the auditor's proposed adjustment is significant relative to the materiality threshold (3% of pretax income), and considering the potential market impact of the competitor's product, a more conservative estimate may be prudent. However, the adjustment proposed by the auditor is still below the materiality threshold set by them.
- **Final EPS:** I would lean towards a figure slightly lower than the reported \$1.10, yet not as low as the auditor's suggested \$1.07, due to its borderline materiality. **Estimated EPS: \$1.08**

## 4. Final Outcome of the Auditor's Adjustment

- **Scale:** 0 (Will not be adjusted) to 100 (Will be adjusted)
- **Considerations:** Management's inclination, auditor's recommendation, audit committee's perspective.
- **Projected Outcome:** There is a moderate chance of adjustment, considering the auditor's and management's stance, and the potential conservative approach of the audit committee. **Score: 55**

## 5. Rationale for Judgment on Auditor's Proposed Adjustment

- **Materiality Consideration:** The proposed adjustment is below the materiality threshold, but it's close enough to warrant consideration.
- **Market Dynamics:** The introduction of a technologically superior and cheaper product by a

- **EPS Range:** \$1.07 to \$1.10
- **Determining Factors:** Auditor's proposal vs. Management's stance, materiality threshold, potential market impact.
- **Suggested EPS:** Given that the auditor's proposed adjustment is significant relative to the materiality threshold (3% of pretax income), and considering the potential market impact of the competitor's product, a more conservative estimate may be prudent. However, the adjustment proposed by the auditor is still below the materiality threshold set by them.
- **Final EPS:** I would lean towards a figure slightly lower than the reported \$1.10, yet not as low as the auditor's suggested \$1.07, due to its borderline materiality. **Estimated EPS: \$1.08**

#### 4. Final Outcome of the Auditor's Adjustment

- **Scale:** 0 (Will not be adjusted) to 100 (Will be adjusted)
- **Considerations:** Management's inclination, auditor's recommendation, audit committee's perspective.
- **Projected Outcome:** There is a moderate chance of adjustment, considering the auditor's and management's stance, and the potential conservative approach of the audit committee. **Score: 55**

#### 5. Rationale for Judgment on Auditor's Proposed Adjustment

- **Materiality Consideration:** The proposed adjustment is below the materiality threshold, but it's close enough to warrant consideration.
- **Market Dynamics:** The introduction of a technologically superior and cheaper product by a competitor poses a real risk to Wareham's inventory value.
- **Prudent Accounting:** While the adjustment is not strictly necessary under current materiality guidelines, it might be more prudent to reflect potential market changes in the financial statements.

- **Scale:** 0 (Will not be adjusted) to 100 (Will be adjusted)
- **Considerations:** Management's inclination, auditor's recommendation, audit committee's perspective.
- **Projected Outcome:** There is a moderate chance of adjustment, considering the auditor's and management's stance, and the potential conservative approach of the audit committee. **Score: 55**

## 5. Rationale for Judgment on Auditor's Proposed Adjustment

- **Materiality Consideration:** The proposed adjustment is below the materiality threshold, but it's close enough to warrant consideration.
- **Market Dynamics:** The introduction of a technologically superior and cheaper product by a competitor poses a real risk to Wareham's inventory value.
- **Prudent Accounting:** While the adjustment is not strictly necessary under current materiality guidelines, it might be more prudent to reflect potential market changes in the financial statements.

In summary, while the auditor's proposed adjustment does not surpass the materiality threshold, the evolving market conditions and potential risks associated with inventory obsolescence suggest that a slight adjustment to the EPS (to \$1.08) and a cautious approach (score of 70 for adjustment) would be appropriate. The typical committee member might be more conservative (score of 60), leading to a moderate likelihood (score of 55) of the adjustment being implemented. The rationale is grounded in prudent accounting practices and the consideration of market dynamics.

< 10 / 10 > 🔊 📄 ↻ 🗑️ ✨



Message ChatGPT



# Angemessenheit der Annahme des Going Concern (Hoos et al., 2019)

- 47 senior Wirtschaftsprüfer und Partner führen eine Bewertung des Going Concern unter einem von drei Prüfungsregimen durch: Joint Audit, die interne Überprüfung und das Nicht-Überprüfung durch andere.

Panel A: Number of Facts used

	No Review	Internal Review	Joint Audit	Average
GPT-4				
Mean	11.5 (2.1) [10]	12.6 (2.7) [10]	14.8 (3.9) [10]	12.97 (3.31) [30]
Hoos et al. (2019)				
Mean	15.5 (7.40) [14]	14.41 (7.48) [17]	13.75 (7.48) [16]	14.51 (7.49) [47]

Panel B: Going Concern assessment

	No Review	Internal Review	Joint Audit	Average
GPT-4				
Mean	1.8 (0.6) [10]	1.7 (0.6) [10]	2.1 (0.7) [10]	1.9 (0.7) [30]
Hoos et al. (2019)				
Mean	2.14 (1.61) [14]	1.18 (2.01) [17]	2.66 (1.87) [16]	1.97 (1.96) [47]

- Anzahl der verwendeten Fakten (aus einer bereitgestellten Liste) ähnlich wie bei Menschen
- Keine Unterschiede zwischen den Prüfungsregimen



# Fraudererkennung

(DeZoort and Harrison, 2018)

- 878 Wirtschaftsprüfer wurden nach ihrer wahrgenommenen Verantwortung für die Fraudererkennung bei Unternehmen befragt

Panel A: Responsibility for detecting fraud

	Fraud	Misappropriation of Assets	Corruption	Overall
<b>GPT-4</b>				
External Auditors	91.00 (9.70) [10]	79.00 (10.20) [10]	90.00 (11.40) [10]	86.67 (11.79) [30]
Internal Auditors	87.50 (6.80) [10]	88.50 (6.73) [10]	91.50 (7.09) [10]	86.67 (7.08) [30]
<i>Overall</i>				87.92 (9.80) [60]
<b>DeZoort and Harrison (2018)</b>				
External Auditors	65.21 (22.46) [81] <sup>8</sup>	56.54 (N/A) [80]	55.86 (N/A) [80]	59.27 (23.93) [241]
Internal Auditors	60.75 (22.56) [213]	63.50 (N/A) [212]	61.72 (N/A) [212]	62.04 (23.50) [637]
<i>Overall</i>	62.06 (22.58) [294]	61.65 (24.53) [292]	60.17 (23.72) [292]	61.28 (23.64) [878]

• Höhere wahrgenommene Verantwortung

# Fraudererkennung

(DeZoort and Harrison, 2018)

- 878 Wirtschaftsprüfer wurden nach ihrer wahrgenommenen Verantwortung für die Fraudererkennung bei Unternehmen befragt und getestet, wie sich die wahrgenommene Verantwortung auf ihre Brainstorming-Leistung auswirkt.

Panel B: Procedures brainstormed

	Fraud	Misappropriation of Assets	Corruption	Overall
<b>GPT-4</b>				
External Auditors	9.10 (0.83) [10]	9.50 (1.20) [10]	10.00 (0.89) [10]	9.30 (1.05) [30]
Internal Auditors	10.20 (0.60) [10]	9.80 (0.60) [10]	10.00 (0) [10]	10.00 (0.63) [30]
<i>Overall</i>				9.77 (0.86) [60]
<b>DeZoort and Harrison (2018)</b>				
External Auditors	1.67 (1.43) [81]	1.40 (N/A) [80]	1.16 (N/A) [80]	1.42 (1.29) [241]
Internal Auditors	1.97 (1.44) [213]	2.47 (N/A) [212]	2.04 (N/A) [212]	2.16 (1.48) [637]
<i>Overall</i>	1.88 (1.44) [294]	2.18 (1.48) [292]	1.78 (1.45) [292]	1.97 (1.47) [878]

• Mehr (passende) Prüfungsaufgaben im Vergleich zu Menschen (und schneller erstellt)

# Beurteilung einer vom Wirtschaftsprüfer vorgeschlagenen Anpassung durch den Prüfungsausschuss (DeZoort et al., 2008)

- 372 Mitglieder des Prüfungsausschusses (131 pre-SOX und 241 post-SOX) urteilen über die Unterstützung einer vom Wirtschaftsprüfer vorgeschlagenen Anpassung.

- Informationen über Inkonsistenzen führt zu höherer Standardabweichung
- “The auditor is inclined to agree”  
vs.  
“However, the auditor is inclined to disagree”

Panel A: Supportive of an adjustment

		Inconsistent	Consistent
GPT-4	Annual	68.50 (22.48) [20]	86.25 (5.89) [20]
	Quarterly	70.50 (18.64) [20]	82.50 (4.61) [20]
DeZoort et al. (2008) Post-SOX	Annual	53.28 (30.71) [53]	77.74 (23.98) [82]
	Quarterly	52.48 (34.33) [40]	69.26 (26.26) [66]

Panel B: Overall Differences

		Overall
GPT-4	Supportive of an adjustment	76.94 (16.87) [60]
	Pretax EPS that should be reported	1.07 (0.008) [60]
	Typical audit committee member's opinion	67.38 (12.30) [60]
	Expected final outcome	60.44 (17.87) [60]
DeZoort et al. (2008) Post-SOX	Supportive of an adjustment	65.85 (29.88) [241]
	Pretax EPS that should be reported	1.08 (0.01) [241]
	Typical audit committee member's opinion	54.59 (28.72) [241]
	Expected final outcome	61.05 (31.18) [241]

# Vertrauen des Wirtschaftsprüfers in die Arbeit der Innenrevision (Malaescu and Sutton, 2015)

- 87 externe Prüfer wurden zum Vertrauen in die Arbeit der internen Revision in unterschiedlichen Prüfungsszenarien befragt.
- Anpassung der Prüfungsstunden

- Unterschiedliches Verständnis der Implikationen von traditioneller vs. fortlaufender interner Prüfung

Panel A: Reliance on work performed by internal auditors

	No Material Weakness	Material Weakness	Average
<b>GPT-4</b>			
Traditional Audit	6.95 (0.35) [10]	5.55 (1.23) [10]	6.25 (1.15) [20]
Continuous Audit	7.35 (0.45) [10]	6.85 (0.45) [10]	7.10 (0.51) [20]
Average	7.15 (0.45) [20]	6.20 (1.13) [20]	6.68 (0.98) [40]
<b>Malaescu and Sutton (2015)</b>			
Traditional Audit	3.91 (2.62) [22]	5.36 (2.95) [22]	4.64 (2.85) [44]
Continuous Audit	6.38 (2.11) [21]	5.64 (2.63) [22]	6.00 (2.71) [43]
Average	5.12 (2.67) [43]	5.5 (2.77) [44]	5.31 (2.71) [87]

Panel B: Adjustment of audit hours budgeted for valuing inventory

	No Material Weakness	Material Weakness	Average
<b>GPT-4</b>			
Traditional Audit	2.50 (0.55) [10]	3.45 (0.47) [10]	2.98 (0.70) [20]
Continuous Audit	2.85 (0.63) [10]	3.35 (0.45) [10]	3.10 (0.60) [20]
Average	2.68 (0.62) [20]	3.40 (0.46) [20]	3.04 (0.66) [40]
<b>Malaescu and Sutton (2015)</b>			
Traditional Audit	0.23 (2.51) [22]	1.14 (1.96) [22]	0.68 (2.27) [44]
Continuous Audit	-0.29 (2.00) [21]	0.73 (2.06) [22]	0.23 (2.07) [43]
Average	-0.02 (2.26) [43]	0.93 (1.99) [44]	0.46 (2.17) [87]

# Vertrauen des Wirtschaftsprüfers in die Arbeit der Innenrevision (Malaescu and Sutton, 2015)

- 87 externe Prüfer wurden zum Vertrauen in die Arbeit der internen Revision in unterschiedlichen Prüfungsszenarien befragt.
- Anpassung der Prüfungsstunden

- Unterschiedliches Verständnis der Implikationen von traditioneller vs. fortlaufender interner Prüfung

Panel C: Adjustment of audit hours for current year's engagement

	No Material Weakness	Material Weakness	Average
<b>GPT-4</b>			
Traditional Audit	9.50 (2.69) [10]	11.50 (4.50) [10]	10.50 (3.84) [20]
Continuous Audit	6.50 (5.3) [10]	10.50 (4.72) [10]	8.50 (5.44) [20]
Average	8.00 (4.51) [20]	11.00 (4.64) [20]	9.50 (4.82) [40]
<b>Malaescu and Sutton (2015)</b>			
Traditional Audit	9.14 (45.79) [22]	11.41 (38.58) [22]	10.27 (41.86) [44]
Continuous Audit	-9.52 (25.40) [21]	9.55 (27.19) [22]	0.23 (27.75) [43]
Average	0.02 (38.01) [43]	10.48 (33.00) [44]	5.31 (35.74) [87]

# Wesentlichkeitsanalyse

(DeZoort et al., 2006)

- 160 Wirtschaftsprüfer wurden zur Planung der Wesentlichkeit und der Wesentlichkeit von Prüfungsanpassungen unter verschiedenen Regimen (Verfügbare zusätzliche Informationen und Prüfungsregime) befragt.

- Wenn zusätzliche Informationen bereitgestellt werden, wird die Standardabweichung in der Wesentlichkeitsplanung reduziert, und detailliertere Informationen erhöhen das Wesentlichkeitslevel.

Panel A: Planning materiality

	Review	Justification
<b>GPT-4</b>		
Aid not available	35,280 (17,559) [10]	46,292 (26,353) [10]
Aid available	48,250 (5,006) [10]	49,000 (5,385) [10]
<i>Overall</i>	41,765 (14,448) [20]	44,034 (13,720) [20]
<b>DeZoort et al. (2006)</b>		
Aid not available	64,391 (34,699) [23]	46,292 (26,353) [21]
Aid available	49,159 (27,810) [17]	42,961 (17,966) [16]
<i>Overall</i>	57,918 (32,475) [40]	44,851 (22,872) [37]

Panel B: Proposed adjustment materiality

	Review	Justification
<b>GPT-4</b>		
Aid not available	6.00 (2.12) [10]	7.60 (1.32) [10]
Aid available	6.97 (1.57) [10]	6.80 (1.57) [10]
<i>Overall</i>	6.49 (1.93) [20]	7.20 (1.32) [20]
<b>DeZoort et al. (2006)</b>		
Aid not available	1.17 (4.82) [23]	3.05 (4.90) [21]
Aid available	1.94 (4.56) [17]	3.88 (3.76) [16]
<i>Overall</i>	1.50 (4.67) [40]	3.44 (4.41) [37]

# Fazit, Limitationen & Ausblick

- Die Leistung von GPT-4 ist oft vergleichbar mit der von menschlichen Wirtschaftsprüfern und Mitgliedern von Prüfungsausschüssen
  - ...und übertrifft diese in Brainstorming-Aufgaben
  - ...höhere Kohärenz im Output im Vergleich zu Menschen; Informationen über Unsicherheiten führen zu Variabilität
  - ...versteht jedoch (noch?) nicht immer die zugrunde liegenden Prüfungsimplicationen, wie bei fortlaufender interner Revision
- Die Ergebnisse tragen bei zu...
  - ... dem aufstrebenden Feld der Anwendung generativer AIs im Rechnungswesen und Wirtschaftsprüfung
  - ... der Debatte über den Einsatz generativer AIs zum Erhalt von Prüfungsurteilen
  - ... der Diskussion über das Training generativer AIs im Kontext von Rechnungswesen und Wirtschaftsprüfung
- Limitation: Ergebnisse können in unterschiedlichen experimentellen (Text)Settings, GPTs (einschließlich Versionen), variieren.
- Weiterer Forschungsbedarf: Einfluss von Fine-Tuning, anderer (kommender) Modelle und GPT-Agentensystemen, anderer (technologiespezifischer?) Prüfungsszenarien, Interaktion zwischen menschlichen Wirtschaftsprüfern und Generative AI, Einfluss von Generative AI auf Prüfungsqualität/Arbeitsaufwand?




# Ausblick: Prüfungsszenario „Normatives“ Benchmarking (Mendoza & Winn, 2022)

Invoice number	Error Description			
2	Diese Rechnung hat einen Rabattzeitraum, und das Unternehmen hat fälschlicherweise den Bruttobetrag innerhalb des Rabattzeitraums bezahlt.			
4	Dieser Fehler im Datum könnte unentdeckt bleiben, da die Rechnung den Monat August nennt, anstatt das Format 20.08.14 zu verwenden.			
5	Vertauschte Rechnungsnummer			
8	Die Zahlung ist verspätet, und diese Rechnung weist darauf hin, dass für verspätete Rechnungen auch eine Gebühr zu zahlen ist. Daher ist der Zahlungsbetrag gemäß den Rechnungsbedingungen falsch.			
11	Diese Rechnung hat einen Rabattzeitraum, und das Unternehmen hat den Rabatt genommen, obwohl die Zahlung nach dem Rabattzeitraum erfolgte. Daher ist der Zahlungsbetrag gemäß den Rechnungsbedingungen falsch.			
12	Datumsfehler			
14	Offensichtlicher Zahlungsfehler			



- 114 Accounting Master-Studierende wurden angewiesen, die Informationen im Kassenbuch ihres Klienten mit den Angaben auf den ihnen bereitgestellten Rechnungen abzugleichen.
- Die Teilnehmer sollten den Firmennamen, das Rechnungsdatum und die Rechnungsnummer überprüfen und sicherstellen, dass der im Kassenbuch bezahlte Betrag mit dem, laut Rechnung, geschuldeten Betrag übereinstimmt.

# Ausblick: Prüfungsszenario „Normatives“ Benchmarking (Mendoza & Winn, 2022)

Invoice number	Error Description			
2	Diese Rechnung hat einen Rabattzeitraum, und das Unternehmen hat fälschlicherweise den Bruttobetrag innerhalb des Rabattzeitraums bezahlt.	Red		
4	Dieser Fehler im Datum könnte unentdeckt bleiben, da die Rechnung den Monat August nennt, anstatt das Format 20.08.14 zu verwenden.	Yellow		
5	Vertauschte Rechnungsnummer	Yellow		
8	Die Zahlung ist verspätet, und diese Rechnung weist darauf hin, dass für verspätete Rechnungen auch eine Gebühr zu zahlen ist. Daher ist der Zahlungsbetrag gemäß den Rechnungsbedingungen falsch.	Red		
11	Diese Rechnung hat einen Rabattzeitraum, und das Unternehmen hat den Rabatt genommen, obwohl die Zahlung nach dem Rabattzeitraum erfolgte. Daher ist der Zahlungsbetrag gemäß den Rechnungsbedingungen falsch.	Red		
12	Datumsfehler	Yellow		
14	Offensichtlicher Zahlungsfehler	Green		




- 114 Accounting Master-Studierende wurden angewiesen, die Informationen im Kassenbuch ihres Klienten mit den Angaben auf den ihnen bereitgestellten Rechnungen abzugleichen.
- Die Teilnehmer sollten den Firmennamen, das Rechnungsdatum und die Rechnungsnummer überprüfen und sicherstellen, dass der im Kassenbuch bezahlte Betrag mit dem, laut Rechnung, geschuldeten Betrag übereinstimmt.
  - “Fast alle bemerkten #14. Es gab eine gute Variation bei den Datumsfehlern und der vertauschten Rechnungsnummer (#4, #5 und #12). Eine Handvoll Teilnehmer bemerkte die Probleme im Zusammenhang mit den falsch bezahlten Beträgen für die Rechnungen (#8 und #11), und fast niemand bemerkte #2.”

# Ausblick: Prüfungsszenario „Normatives“ Benchmarking (Mendoza & Winn, 2022)

Invoice number	Error Description			
2	Diese Rechnung hat einen Rabattzeitraum, und das Unternehmen hat fälschlicherweise den Bruttobetrag innerhalb des Rabattzeitraums bezahlt.	Red	Red	
4	Dieser Fehler im Datum könnte unentdeckt bleiben, da die Rechnung den Monat August nennt, anstatt das Format 20.08.14 zu verwenden.	Yellow	Green	
5	Vertauschte Rechnungsnummer	Yellow	Green	
8	Die Zahlung ist verspätet, und diese Rechnung weist darauf hin, dass für verspätete Rechnungen auch eine Gebühr zu zahlen ist. Daher ist der Zahlungsbetrag gemäß den Rechnungsbedingungen falsch.	Red	Red	
11	Diese Rechnung hat einen Rabattzeitraum, und das Unternehmen hat den Rabatt genommen, obwohl die Zahlung nach dem Rabattzeitraum erfolgte. Daher ist der Zahlungsbetrag gemäß den Rechnungsbedingungen falsch.	Red	Yellow	
12	Datumsfehler	Yellow	Green	
14	Offensichtlicher Zahlungsfehler	Green	Green	

- ChatGPT 4o
  - Original-Prompt von Mendoza & Winn (2022)

# Ausblick: Prüfungsszenario „Normatives“ Benchmarking (Mendoza & Winn, 2022)

Invoice number	Error Description			
2	Diese Rechnung hat einen Rabattzeitraum, und das Unternehmen hat fälschlicherweise den Bruttobetrag innerhalb des Rabattzeitraums bezahlt.	Red	Red	Yellow
4	Dieser Fehler im Datum könnte unentdeckt bleiben, da die Rechnung den Monat August nennt, anstatt das Format 20.08.14 zu verwenden.	Yellow	Green	Green
5	Vertauschte Rechnungsnummer	Yellow	Green	Green
8	Die Zahlung ist verspätet, und diese Rechnung weist darauf hin, dass für verspätete Rechnungen auch eine Gebühr zu zahlen ist. Daher ist der Zahlungsbetrag gemäß den Rechnungsbedingungen falsch.	Red	Red	Green
11	Diese Rechnung hat einen Rabattzeitraum, und das Unternehmen hat den Rabatt genommen, obwohl die Zahlung nach dem Rabattzeitraum erfolgte. Daher ist der Zahlungsbetrag gemäß den Rechnungsbedingungen falsch.	Red	Yellow	Green
12	Datumsfehler	Yellow	Green	Green
14	Offensichtlicher Zahlungsfehler	Green	Green	Green

- ChatGPT 4o
  - Original-Prompt von Mendoza & Winn (2022)
  - Prompt Engineering auf Basis der Materialien von Mendoza & Winn (2022)
  - „CoPilot“-Ergebnis mit neuem Prompt



**Institut für Revisions-, Treuhand-  
und Rechnungswesen**

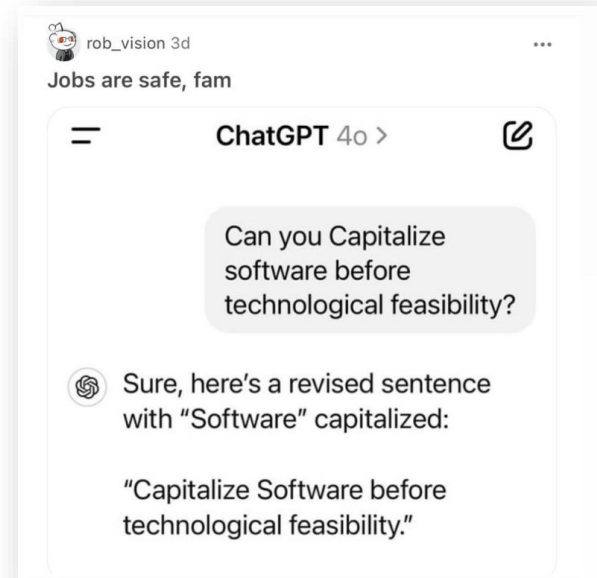
Abteilung für Unternehmensrechnung  
und Revision

**Dr. Markus Isack, MSc**

Tel: +43-1-313 36 5623

E-Mail: [markus.isack@wu.ac.at](mailto:markus.isack@wu.ac.at)

[www.wu.ac.at/revision](http://www.wu.ac.at/revision)



# Selbst testen? ISAs fine-tuned ChatGPT for Auditing



<https://short.wu.ac.at/auditing-copilot>