

## Inhaltsverzeichnis

1. Einleitung .....	1
2. Simulationsbasiertes Lernen .....	2
2.1. Anleitung der Studierenden .....	3
2.2. Darstellungsformat der Fallsimulationen .....	3
2.3. Feedback für die Studierenden .....	3
2.4. Kooperatives simulationsbasiertes Lernen .....	4
2.5. Fragestellungen .....	4
3. Künstliche Intelligenz in Lernumgebungen.....	4
3.1. Sprachliche Analyse von Antworttexten .....	5
3.2. Strukturelle Analyse von Antworttexten .....	6
3.3. Inhaltlich-strukturelle Analyse von Antworttexten .....	6
3.4. Fragestellungen .....	7
4. FAMULUS: Simulationsbasiertes Lernen mit adaptivem automatischem Feedback im Medizin- und Lehramtsstudium .....	7
5. Zusammenfassung und Ausblick .....	8
Literaturverzeichnis.....	9
Autoren.....	11

### 1. Einleitung

Die Vermittlung theoretischen Wissens zählt zum Kerngeschäft von Universitäten, der Aufbau und die Förderung eines konzeptionellen Vorstellungsvermögens von Studierenden ist entsprechend ein zentrales Anliegen der akademischen Lehre. Häufig wird der praktischen Anwendung von Wissen jedoch weniger Aufmerksamkeit geschenkt. Möglichkeiten zum eigenen Handeln und Ausprobieren in beruflichen Anwendungssituationen mit entsprechendem lernförderlichem Feedback sind nur selten gegeben. Da problemlösendes Denken die Lücke zwischen dem vermittelten Wissen und der Anwendung in der Praxis schließen kann (Funke, 2003), ist der Aufbau von *Problemlösekompetenzen* eine wichtige Forderung im Curriculum vieler Studiengänge, etwa in der Medizin (MFT, 2015) und Informatik (The Joint Task Force on Computing Curricula, 2013).

Zur Förderung von Problemlösekompetenzen bieten sich Fallsimulationen an, in denen ein problemorientiertes Szenario geschildert wird, da Studierende im Rahmen dieser Simulationen eigenständig eine Lösung finden müssen (Ziv, Wolpe, Small & Glick, 2003).

Um erfolgreiches Lernen zu gewährleisten, sollten die Studierenden zudem individuelles Feedback bezüglich ihrer Lösungsansätze erhalten.

In diesem Beitrag wird ein Überblick über den derzeitigen Wissens- und Forschungsstand zum erfolgreichen Einsatz von Fallsimulationen sowie zu Methoden der Künstlichen Intelligenz für die automatische Generierung von adaptivem Feedback gegeben. Im Anschluss werden Fragestellungen diskutiert, die sich durch die Kombination dieser zwei Unterstützungsmöglichkeiten zum Lernen ergeben. Insbesondere stellen wir das BMBF-geförderte Projekt *FAMULUS* vor, das sich mit dem Einsatz von computergestützten Fallsimulationen mit automatischem Feedback zur Förderung von Diagnosekompetenzen in der Medizin- und Lehramtsausbildung befasst.

## 2. Simulationsbasiertes Lernen

Simulationen ermöglichen sowohl authentische Darstellungen von relevanten Problemen aus der Berufspraxis als auch interaktive Bearbeitungsformen dieser Probleme. Durch die Bearbeitung von realistischen Problemen liegt der Fokus von Simulationen folglich auf der Anwendung von Wissen in komplexen beruflichen Situationen. Darüber hinaus können Lernende innerhalb von Simulationen die vorliegende Situation durch ihr eigenes Handeln beeinflussen. Hierdurch werden hohe Anteile konstruktiver und interaktiver Lernaktivitäten ermöglicht. Lernaktivitäten, die einen kreativen Umgang mit den gegebenen Materialien (konstruktive Aktivitäten) und eine gut strukturierte Interaktion mit anderen Lernenden (interaktive Aktivitäten) ermöglichen, gelten als besonders wichtig für den Lernerfolg beim Einsatz digitaler Medien, und somit auch beim Einsatz von Simulationen (Stegmann & Fischer, 2016).

Eine Fallsimulation beginnt in der Regel zunächst mit der Beschreibung einer Problemsituation. Im Kontext der Medizin könnte dies beispielsweise ein Patient sein, der unter bestimmten Symptomen leidet. Im Folgenden versucht der Lernende basierend auf den gegebenen Informationen herauszufinden, welche Ursachen dem Problem zu Grunde liegen und wie weiter vorgegangen werden sollte. In einer medizinischen Fallsimulation könnte ein Lernender beispielsweise Untersuchungen einleiten oder Fragen stellen, um herauszufinden, was dem Patienten fehlt und wie er behandelt werden sollte. Am Ende dieses interaktiven Problemlöseprozesses wird der Lernende in der Regel dazu aufgefordert, eine Lösung für das Problem aufzuzeigen. Im Bereich Medizin muss beispielsweise eine Diagnose gestellt (bzw. ausgewählt) und ein weiteres Vorgehen vorgeschlagen (bzw. ausgewählt) werden.

In der Medizinischen Ausbildung werden computerbasierte Fallsimulationen bereits seit Ende der 1990er Jahre mit großem Erfolg eingesetzt (Fischer, 2000). Auch in anderen Bereichen, wie beispielsweise der Lehramtsausbildung, werden computerbasierte Fallsimulationen vermehrt angewendet (z.B. Kollar et al. 2014). Forschungsarbeiten auf diesem Gebiet versuchen vor allem herauszufinden, *wie* computerbasierte Fallsimulationen den bestmöglichen Lerneffekt für die Studierenden erzielen. Die bisherige Forschung zum computergestützten fallbasierten Lernen konzentrierte sich auf (1) die Anleitung der Studierenden während der Fallsimulation, (2) das Darstellungsformat der Fälle und (3) Feedback für die Studierenden. Darüber hinaus wird (4) die Sozialform des Lernens als weiteres Forschungsfeld thematisiert.

## 2.1. Anleitung der Studierenden

In der Lehr-Lernforschung besteht weitgehend Konsens, dass für das Lernen mittels anspruchsvoller Problemstellungen eine zusätzliche instruktionale Unterstützung erforderlich ist (Kirschner et al., 2006). Kurze Unterbrechungen der Lernaktivität durch spezifische Aufgaben- oder Fragestellungen bieten eine Möglichkeit solch eine Unterstützung in Simulationen zu integrieren. Sie können die Aufmerksamkeit des Lernenden auf wichtige Aktivitäten innerhalb des Lernprozesses lenken (Quintana et al., 2004). Besonders geeignet für Fallsimulationen sind *Selbsterklärungsprompts*, bei denen Lernende aufgefordert werden, eine spezifische Art der Zusammenfassung der einzelnen bearbeiteten Schritte innerhalb der Fallsimulation anzufertigen. Studien zeigen, dass solche Prompts zu einer qualitativ besseren Problemlösung führen (Chamberland et al., 2011; 2015). Zurzeit gibt es eine intensive Diskussion, *wie genau* die Unterbrechung gestaltet und in welcher Phase des Lernens sie stattfinden sollte, um optimale Lerneffekte zu erzielen (Kalyuga, 2007).

## 2.2. Darstellungsformat der Fallsimulationen

Weiterhin wurde viel spekuliert, bisher jedoch wenig empirisch überprüft, *welches* Darstellungsformat für das Lernen in Fallsimulationen am besten geeignet ist und wovon dies abhängt. Eine Möglichkeit ist, den Studierenden alle Informationen über einen Fall direkt am Anfang zur Verfügung zu stellen – also im Kontext der Medizin alle Symptome des Patienten, alle Befunde und alle Tests auf einmal darzustellen. Dies wird als das sogenannte *Whole-Case Format* (holistisches Darstellungsformat) bezeichnet. Alternativ gibt es die Möglichkeit, den Studierenden zu Anfang nur wenige Informationen über den Patienten zu geben und die Studierenden selbst auswählen zu lassen, welche Informationen sie sich als nächstes ansehen möchten – das sogenannte *Serial-Cue Format* (serielles Darstellungsformat). Einige Forscher sind der Meinung, dass das serielle Darstellungsformat Studierende mit wenig Vorwissen überfordert, weil diese dadurch zu früh in Entscheidungssituationen gebracht werden, denen sie nicht gewachsen sind (z.B. die Entscheidung darüber, welcher medizinische Test als erstes angewendet werden soll) (Schmidt & Mamede, 2015). Fortgeschrittene Studierende könnten hingegen von der seriellen Darstellung profitieren, da sie bereits über funktionale Schemata verfügen und somit zielgerichtet, ohne Überlastung des Arbeitsgedächtnisses, vorgehen können. Generell könnte eine holistische Darstellungsform aber auch für Studierende mit wenig Vorwissen überfordernd sein, da sie die Menge an Informationen nicht bewältigen können und ihr Arbeitsgedächtnis bei der Bearbeitung überlastet wird (Sweller, 1988).

## 2.3. Feedback für die Studierenden

Eine weitere offene Forschungsfrage ist, *wie* Feedback für die Studierenden am effektivsten in Fallsimulationen eingesetzt werden kann. Heitzmann (2014) konnte zeigen, dass Feedback bezüglich der Falllösungen die Leistung der Studierenden positiv beeinflusst. Allerdings verschwand dieser positive Effekt, wenn neben Feedback zu den Falllösungen auch die oben beschriebenen Selbsterklärungsprompts eingesetzt wurden. Eine Idee zur Erklärung dieser Ergebnisse ist, dass das Feedback bisher nur auf den

Inhalt der Lösung, nicht aber auf die Richtigkeit der Argumentationsstruktur und des schlussfolgernden Denkens der Studierenden einging. Individuelles Feedback durch einen Experten ist eine Möglichkeit, auf das schlussfolgernde Denken der Studierenden gezielter einzugehen. Da dies in der Hochschulpraxis allerdings einen hohen Aufwand erfordert und in der Regel nicht zeitnah möglich ist, ist dieser Weg häufig nur schwer umsetzbar. Aus diesen (zeit-)ökonomischen Gründen wird in der Hochschulpraxis zumeist auf *statisches Expertenfeedback* zurückgegriffen, welches aus der Bereitstellung einer von Experten angefertigten Musterlösung besteht. Bei diesen statischen Feedbackformen liegt es an den Studierenden, die eigene Lösung mit der Musterlösung zu vergleichen. Ein *automatisches adaptives Feedbacksystem* könnte hier Abhilfe schaffen, da für den Experten während der Feedbackgabe kein Aufwand anfällt und sich das Feedback adaptiv auf die Lösung der Studierenden bezieht.

## 2.4. Kooperatives simulationsbasiertes Lernen

Während sich die bisherigen Forschungsfragen zu computergestützten Fallsimulationen mit *individuellem* Lernen auseinandersetzen, ist der tatsächliche Alltag häufig durch *kooperativ* zu lösende Aufgaben gekennzeichnet. Hautz et al. (2015) konnten zeigen, dass Studierende, die gemeinsam an einer Problemlösung arbeiten, die Leistungen individueller Studierender übertreffen. Abgesehen von dieser Studie wird, häufig wegen des großen organisatorischen Aufwands, auf kooperative Settings verzichtet, sodass über die Effekte kooperativen Problemlösens bisher wenig bekannt ist (Kiesewetter, Fischer & Fischer, 2017).

## 2.5. Fragestellungen

Zusammenfassend kann man sagen, dass bisher auf keinem der vier vorgestellten Forschungsgebiete zu Fallsimulationen Konsens darüber herrscht, welche Methodiken den bestmöglichen Lerneffekt erzielen. Wir halten hier insbesondere drei offene Fragestellungen fest:

- 1) Welches Darstellungsformat (holistisch oder seriell) ist am besten geeignet?
- 2) Welche Effekte hat adaptives automatisches Feedback auf Lernende im Vergleich zu statischem Expertenfeedback?
- 3) Wie wirkt sich kooperatives Lernen, verglichen mit individuellem Lernen, in Fallsimulationen aus?

Um die erste und letzte Fragestellung beantworten zu können, sind vor allem didaktische Studien notwendig. Zur Beantwortung der zweiten Fragestellung hingegen, muss zudem ein automatisches Feedback System für die jeweilige Fallsimulation vorhanden sein. Im Folgenden geben wir daher einen Überblick über den derzeitigen Forschungsstand zur automatischen Generierung von adaptivem Feedback durch Methoden aus der Künstlichen Intelligenz.

## 3. Künstliche Intelligenz in Lernumgebungen

In den vergangenen zwanzig Jahren hat sich die Forschung im Bereich Künstliche Intelligenz rapide entwickelt und dadurch völlig neue Anwendungsgebiete erschlossen. Die

Hochschulbildung ist eines dieser Anwendungsgebiete und könnte in Zukunft enorm von intelligenten Systemen profitieren (Luckin, Holmes, Griffiths & Forcier, 2016). In Übereinstimmung mit der oben formulierten Forderung, nicht nur Wissen zu vermitteln, sondern auch Problemlösekompetenzen zu fördern, werden zunehmend intelligente Systeme entwickelt, die nicht nur domänenspezifisches Wissen vermitteln, sondern den Fokus auf geeignete Unterstützungsmaßnahmen beim Bearbeiten komplexer Problemstellungen legen (Roll & Wylie, 2016). Solche Systeme werden *intelligente Tutoren* genannt. Sie erzeugen automatisch passendes Feedback während einer Lernaktivität und ermöglichen so eine individuelle Betreuung und Förderung der Lernenden, was in der Hochschullehre wegen Personal- und Zeitmangel oft kaum möglich ist. Neben der adaptiven Strukturierung von Lerneinheiten, wie zum Beispiel in AutoTutor (Graesser, Wiemer-Hastings, Wiemer-Hastings & Kreuz, 1999), eTeacher (Schiaffino, Garcia & Amandi, 2008) und ZOSMAT (Keleş, Ocağ, Keleş & Gülcü, 2009), spielt die *automatische und individuelle Einschätzung von Lösungen*, die Studierende zu einer gegebenen Problemstellung erarbeitet haben, eine entscheidende Rolle.

Zu geschlossenen Aufgabenformaten, wie Multiple-Choice-Tests oder einfachen Rechenaufgaben, kann eine Bewertung und Feedback zu Lösungen Studierender leicht vordefiniert und dann unverändert angezeigt werden. Gerade bei komplexeren Problemlöseaufgaben eignen sich dagegen offene Antwortformate, beispielsweise in Essayform, da hierdurch Lernprozesse und Begründungen von Studierenden sichtbar gemacht werden können. Analysen solcher Lernprozesse können bei der Suche nach Erklärungen für die (fehlende) Wirksamkeit instruktorischer Maßnahmen von besonderer Wichtigkeit sein (Wecker, Fischer & Stegmann, 2012). Die Bewertung von Texten aus offenen Antwortformaten ist deutlich herausfordernder als die Bewertung geschlossener Antwortformate, da dies die Integration von Methoden der automatischen Sprachanalyse aus dem Bereich *Natural Language Processing* (NLP) in intelligente Tutoren-Systeme erfordert.

### 3.1. Sprachliche Analyse von Antworttexten

Für Oberflächenmerkmale wie die durchschnittliche Satzlänge, die Vielfalt der Wortwahl oder die Analyse der Satzstruktur liegen bereits zahlreiche NLP-Methodensammlungen vor, z.B. DKPro Core (Eckart de Castilho & Gurevych, 2014), Stanford CoreNLP (Manning et al., 2014) oder NLTK (Bird, Klein & Loper, 2009), die zur Bewertung von Antworttexten eingesetzt werden können. Fehler bezüglich dieser Aspekte, z.B. eine inkorrekte Satzstruktur oder die häufige Verwendung desselben Wortes, können automatisch identifiziert werden, indem die Werte im zu bewertenden Text mit Richtwerten verglichen werden. Diese Richtwerte werden durch die Analyse einer Großzahl von als gut eingestufte Beispiel-Antworttexte ermittelt. Für die durchschnittliche Satzlänge könnte ein solcher Richtwert zum Beispiel zwischen 10 und 25 Wörtern pro Satz vorsehen. Finden sich in Antworttexten Studierender Sätze, deren automatisch ermittelte Länge außerhalb dieses Intervalls liegt, erhalten die Studierenden Feedback mit dem Vorschlag, die Satzlänge entsprechend anzupassen. In ähnlicher Weise wird die Struktur eines Satzes in Antworttexten Studierender anhand ihrer Vorkommenshäufigkeit in gut geschriebenen Beispiel-Antworttexten bewertet. Kommt die Satzstruktur selten in den Beispieltextrn vor, steigt die Wahrscheinlichkeit, dass die Struktur inkorrekt ist. Diese Ansätze betrach-

ten vor allem den Stil und die Korrektheit der *Sprache*, nicht aber den Aufbau und Inhalt eines Antworttextes.

### 3.2. Strukturelle Analyse von Antworttexten

In den vergangenen Jahren haben es *maschinelle Lernverfahren* („Machine Learning“) ermöglicht auch komplexere Aspekte von Antwort-Essays zu untersuchen, wie zum Beispiel Kohärenz (Higgins, Burstein, Marcu & Gentile, 2004; Miltsakaki & Kukich, 2004; Somasundaran, Burstein & Chodorow, 2014), Argumentationsstruktur (Persing & Ng, 2016; Stab & Gurevych, 2017) und Kausalzusammenhänge (Hastings, Hughes, Britt, Blaum & Wallace, 2014; Hughes, Hastings, Britt, Wallace & Blaum, 2015). Basierend auf diesen Aspekten kann die *Qualität* der Kohärenz und der Essaystruktur eingeschätzt werden. Um Aussagen über die *Kohärenz* eines Essays machen zu können, benötigen maschinelle Lernverfahren hunderte Beispiessays mit hoher und niedriger Kohärenz. Aus diesen Trainingsbeispielen erstellt das automatische Lernverfahren ein internes Modell, anhand dessen die Kohärenz von beliebigen neuen Essays bestimmt werden kann (Liu, Li, Xu & Liu, 2017). Da dieses interne Modell aus Zahlen und mathematischen Formeln besteht, ist es jedoch nur schwer möglich, eine für Menschen verständliche Erklärung zu liefern, die angibt, warum das Verfahren einen bestimmten Kohärenzwert für einen Essay ermittelt hat. Bisher wird daher häufig allgemeines Feedback generiert, welches auf typische Fehler aufmerksam macht. Dies kann Studierende zwar dazu bewegen, über bestimmte Aspekte ihres Essays nachzudenken, aber kaum eine Förderung individueller Lernbedarfe erreichen. Für Essays mit niedriger Kohärenzvorhersage nutzen Liu et al. (2017) beispielsweise indirekte Fragen, wie „Werden Satzüberleitungen in einem Absatz verwendet?“ und „Sind verschiedene Absätze mit Verbindungswörtern versehen?“, als Feedback. Um Aussagen über die Qualität der *Argumentationsstruktur* in einem Essay machen zu können, muss die Argumentationsstruktur in hunderten Beispiessays von Hand gekennzeichnet werden. Anhand dieser Kodierung wird ein maschinelles Lernverfahren für die Bestimmung der argumentativen Struktur in beliebigen neuen Essays trainiert. Um adaptives Feedback bezüglich der ermittelten Struktur eines konkreten Essays geben zu können, erfolgt in einem zweiten Schritt ein Abgleich der ermittelten Struktur mit einer vordefinierten, idealen Argumentationsstruktur. Im Criterion System (Burstein, Chodorow & Leacock, 2004; Burstein, Marcu & Knight, 2003) werden Studierende beispielsweise darauf hingewiesen, wenn ihr Essay keine Schlussfolgerung oder zu wenige argumentative Elemente enthält.

### 3.3. Inhaltlich-strukturelle Analyse von Antworttexten

Die bisher diskutierten Methoden beschränken sich auf die Analyse und Bewertung der Qualität und Richtigkeit der Struktur von Essays – unabhängig vom *Inhalt*. Das bedeutet, die inhaltliche Richtigkeit und ob die Struktur sinnvoll bezüglich des Inhaltes ist, wird zumeist nicht evaluiert, da dies bisher nicht verlässlich automatisch möglich ist. Derzeitige Systeme, die Inhalt und Struktur gemeinsam untersuchen und bewerten, benötigen ein vordefiniertes *Wissens- und Kausalmodell* einer Essaythematik (Green, 2016; Hughes et al., 2015). Dies ist sehr arbeitsaufwendig, da Experten für jede Problemstellung ein neues Modell erarbeiten müssen. Zudem erlaubt es wenig Spielraum bezüglich verschiedener Lösungsvarianten, die nicht mit dem vordefinierten Modell übereinstimmen. Allerdings kann konkreteres adaptives Feedback gegeben werden, das sich sowohl auf

die Struktur als auch auf den Inhalt eines Essays bezieht und beispielsweise aufzeigt, dass in einer Schlussfolgerung ein wichtiger Schritt übersprungen oder relevante Hintergrundinformationen nicht genannt wurden. Hastings et al. (2014) und Hughes et al. (2015) haben kürzlich einen vielversprechenden Ansatz mit einem vordefiniertem Kausalmodell entwickelt, bei dem das Modell alle wichtigen Konzepte abdecken muss, die ein Studierenden-Essay enthalten könnte. Im Kontext der Medizin wären solche Konzepte mögliche Diagnosen, Befunde, und Symptome, die im Kausalmodell beispielsweise durch „führt zu“ (Symptom und Befund) oder „lässt schließen“ (Befund/Symptom und Diagnose) Relationen verbunden sind. Um diese Konzepte in ihren verschiedenen Formulierungen automatisch in einem Essay zu erkennen, werden die Konzepte zunächst in einer Vielzahl von Beispiessays gekennzeichnet. Des Weiteren werden die Kausalzusammenhänge, die in diesen Essays zwischen den Konzepten bestehen, markiert. Mit den gekennzeichneten Essays wird dann ein maschinelles Lernverfahren darauf trainiert, die verschiedenen Konzepte und Kausalzusammenhänge in neuen Essays zu erkennen, selbst wenn die neuen Essays andere Formulierungen als die Beispiessays enthalten. Um darauf aufbauend die inhaltliche Richtigkeit eines Essays zu bewerten, werden die automatisch erkannten Konzepte und die Kausalstruktur mit dem vordefinierten Kausalmodell verglichen. Basierend auf falschen oder nicht vorhandenen Kausalverbindungen und Konzepten im untersuchten Essay kann dann konkretes adaptives Feedback für Studierende generiert werden.

### 3.4. Fragestellungen

Derzeitige intelligente Tutoren-Systeme beschränken sich also häufig auf die Bewertung von Antworttexten ohne oder mit sehr allgemeinem Feedback. Ob und wie es möglich ist effektiv und automatisch adaptives Feedback zu generieren, das auf komplizierte Aspekte wie die Kausalstruktur und inhaltliche Richtigkeit von Begründungen in Antworttexten eingeht, ist eine offene Forschungsfrage.

## 4. FAMULUS: Simulationsbasiertes Lernen mit adaptivem automatischem Feedback im Medizin- und Lehramtsstudium

Das BMBF-geförderte Projekt FAMULUS beschäftigt sich mit den drei zuvor genannten offenen Fragestellungen zur effektiven Gestaltung von Fallsimulationen und der damit verbundenen Frage zur automatischen Generierung von adaptivem Feedback. Konkret werden die folgenden Fragen untersucht:

- 1) Welches Darstellungsformat (holistisch oder seriell) ist am besten für Fallsimulationen geeignet?
- 2) Wie kann adaptives Feedback automatisch generiert werden und welche Auswirkungen hat es auf Lernende im Vergleich zu statischem Expertenfeedback?
- 3) Wie wirkt sich kooperatives Lernen, verglichen mit individuellem Lernen, in Fallsimulationen aus?

Der Fokus liegt dabei auf Fallsimulationen zur Förderung einer spezifischen Art von Problemlösekompetenzen, die für viele Studiengänge relevant und für viele berufliche Tätigkeiten kritisch sind, nämlich Diagnosekompetenzen. Untersucht werden die Effekte von Fallsimulationsformaten und adaptivem Feedback auf Diagnosekompetenzen anhand der zahlenmäßig großen Bereiche des Medizin- und Lehramtsstudiums.

Diagnostizieren wird als das zielgerichtete Sammeln und Integrieren von Informationen zum Zweck der Unsicherheitsreduktion verstanden, um bestmögliche Entscheidungen treffen zu können. So ist es beispielsweise für eine Lehrkraft unabdingbar, verhaltensauffällige oder leistungsschwache Schüler zu identifizieren, um über das passende Vorgehen oder die passende Intervention für den jeweiligen Schüler zu entscheiden. Auch in der Medizin sind Diagnosekompetenzen unentbehrlich, um Problemlösesituationen im professionellen Umfeld erfolgreich bewältigen zu können, wie beispielsweise die Wahl der passenden Therapie für eine Patientin. Auch über diese beiden Bereiche hinaus, wie zum Beispiel in den Rechts- und Wirtschaftswissenschaften oder der Informatik, sind Diagnosekompetenzen von erheblicher Bedeutung in der beruflichen Praxis.

Zur Untersuchung der oben genannten FAMULUS-Fragestellungen werden insgesamt drei empirische Studien, sowohl im Medizin- als auch im Lehramtsstudium, durchgeführt. Parallel dazu wird mit verschiedenen maschinellen Lernverfahren experimentiert, um ein möglichst robustes intelligentes Tutoren-System zu entwickeln, welches Studierenden konkretes automatisch generiertes Feedback zu ihren Diagnosen liefert. Dieses Feedback soll sowohl eine Einschätzung des Inhalts als auch der Denk- und Lösungsstruktur der jeweiligen Diagnose beinhalten.

Die erste FAMULUS-Studie vergleicht das holistische und serielle Fallformat miteinander und überprüft, ob sich der Lernerfolg im jeweiligen Format als vorwissensabhängig erweist. Gleichmaßen sollen Lernprozessanalysen anzeigen, wie sich Studierende mit wenig bzw. viel Vorwissen durch Fälle im seriellen Darstellungsformat bewegen. Während dieser Fallsimulation schreiben Studierende offene Antworttexte, in denen sie eine Diagnose stellen und diese ausführlich begründen. Diese Texte werden im nächsten Schritt von Experten bezüglich ihrer Struktur und ihres Inhalts gekennzeichnet und als Beispiellantworten zum Trainieren maschineller Lernverfahren verwendet. Dadurch wird ein neues intelligentes Tutoren-System entwickelt, das automatisch adaptives Feedback für Antworttexte generiert. In der zweiten FAMULUS-Studie wird die Effektivität dieses automatischen Feedbacks mit der Effektivität eines statischen Expertenfeedbacks, in Form einer zur Verfügung gestellten Musterlösung, experimentell verglichen. Zudem werden die Effekte der verschiedenen Falldarstellungsformate und des automatischen Feedbacks auch im kooperativen Lernsetting überprüft und mit dem individuellen Lernsetting verglichen. In der dritten FAMULUS-Studie werden die verschiedenen Fallformate, das automatische Feedback und das kollaborative im Vergleich zum individuellen Setting im Feldeinsatz an mehreren Hochschulen untersucht.

## 5. Zusammenfassung und Ausblick

Wie beispielsweise die BMBF-Förderlinie „Forschung zur digitalen Hochschulbildung“ zeigt, spielt Digitalisierung eine zunehmende Rolle in der Hochschulbildung. Ein wichtiger Aspekt der digitalen Hochschulbildung sind digitale Lernumgebungen, in denen Studierende, unterstützt durch digitale Systeme, eigenständig lernen. Wie solche Lernumgebungen, und insbesondere Fallsimulationen, am besten gestaltet werden, um deren effektiven Einsatz zur Verbesserung von praktischen Problemlösekompetenzen zu gewährleisten, ist bisher jedoch unklar. Zudem gibt es nur wenige Ansätze, adaptives, detailliertes Feedback automatisch und effektiv als Teil von digitalen Lernumgebungen zu generieren. Das FAMULUS-Projekt untersucht daher offene Fragestellungen bezüglich verschiedenen Formaten der Fallsimulation und der Generierung von automati-



schem adaptivem Feedback in digitalen Lernumgebungen. Ziel des Projekts ist es, dass die gewonnenen Erkenntnisse in Zukunft als didaktische Grundlage zur Gestaltung von Lernumgebungen, und insbesondere Simulationen, dienen können.

Das FAMULUS-Projekt ist somit ein wichtiger Schritt in Richtung des flächendeckenden Einsatzes digitaler Lernumgebungen, die durch authentische Fallsimulationen und individuelles Feedback praxisorientierte Problemlösekompetenzen vermitteln.

## Literaturverzeichnis

- Bird, S., Klein, E. & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media, Inc.
- Burstein, J., Chodorow, M. & Leacock, C. (2004). Automated Essay Evaluation: The Criterion Online Writing Service. *AI Magazine*, 25(3), 27–36.
- Burstein, J., Marcu, D. & Knight, K. (2003). Finding the WRITE Stuff: Automatic Identification of Discourse Structure in Student Essays. *IEEE Intelligent Systems*, 18(1), 32–39.
- Chamberland, M., St-Onge, C., Setrakian, J., Lanthier, L., Bergeron, L., Bourget, A., Mamede, S., Schmidt, H. & Rikers, R. (2011). The influence of medical students' self-explanations on diagnostic performance. *Medical Education*, 45(7), 688–695.
- Chamberland, M., Mamede, S., St-Onge, C., Setrakian, J., Bergeron, L. & Schmidt, H. (2015). Self-explanation in learning clinical reasoning: the added value of examples and prompts. *Medical Education*, 49(2), 193–202.
- Eckart de Castilho, R. & Gurevych, I. (2014). A Broad-Coverage Collection of Portable NLP Components for Building Shareable Analysis Pipelines. In *Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT* (pp. 1–11).
- Fischer, M. R. (2000). CASUS - An authoring and learning tool supporting diagnostic reasoning. *Zeitschrift für Hochschuldidaktik*, 1(1), 87–98.
- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., & Kreuz, R. (1999). AutoTutor: A Simulation of a Human Tutor. *Cognitive Systems Research*, 1(1), 35–51.
- Green, N. L. (2016). Argumentation Scheme-Based Argument Generation to Support Feedback in Educational Argument Modeling Systems. *International Journal of Artificial Intelligence in Education*, 27(3), 515–533.
- Hastings, P., Hughes, S., Britt, A., Blaum, D. & Wallace, P. (2014). Toward Automatic Inference of Causal Structure in Student Essays. In *Proceedings of the 12th International Conference on Intelligent Tutoring Systems (ITS'14)* (pp. 266–271).
- Heitzmann, N. (2014): *Fostering diagnostic competence in different domains*. Dissertation, LMU München: Fakultät für Psychologie und Pädagogik
- Higgins, D., Burstein, J., Marcu, D. & Gentile, C. (2004). Evaluating Multiple Aspects of Coherence in Student Essays. In *Proceedings of the Human Language*

- Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL'04) (pp. 185–192).
- Hautz, W. E., Kämmer, J. E., Schaubert, S. K., Spies, C. D. & Gaissmaier, W. (2015). Diagnostic performance by medical students working individually or in teams. *JAMA*, 313(3), 303–304.
- Hughes, S., Hastings, P., Britt, M. A., Wallace, P. & Blaum, D. (2015). Machine Learning for Holistic Evaluation of Scientific Essays. In Proceedings of the 17th International Conference on Artificial Intelligence in Education (AIED'15) (pp. 165–175).
- Keleş, A., Ocağ, R., Keleş, A. & Gülcü, A. (2009). ZOSMAT: Web-based Intelligent Tutoring System for Teaching-Learning Process. *Expert Systems with Applications*, 36(2), 1229–1239.
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, 19(4), 509–539.
- Kirschner, P. A., Sweller, J. & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86.
- Kollar, I., Ufer, S., Reichersdorfer, E., Vogel, F., Fischer, F. & Reiss, K. (2014). Effects of heuristic worked examples and collaboration scripts on the acquisition of mathematical argumentation skills of teacher students with different levels of prior knowledge. *Learning and Instruction*, 32, 22–36.
- Liu, M., Li, Y., Xu, W. & Liu, L. (2017). Automated Essay Feedback Generation and Its Impact in the Revision. *IEEE Transactions on Learning Technologies*, erscheint.
- Luckin, R., Holmes, W., Griffiths, M. & Forcier, L. B. (2016). *Intelligence Unleashed: An Argument for AI in Education*. Pearson.
- Manning, C. D., Bauer, J., Finkel, J., Bethard, S. J., Surdeanu, M. & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics (ACL'14) (pp. 55–60).
- Miltsakaki, E. & Kukich, K. (2004). Evaluation of Text Coherence for Electronic Essay Scoring Systems. *Natural Language Engineering*, 10(1), 25–55.
- MFT (Medizinischer Fakultätentag der Bundesrepublik Deutschland e. V.) (2015). *Nationaler Kompetenzbasierter Lernzielkatalog Medizin*. 76. Ordentlicher Medizinischer Fakultätentag am 04.06.2015 in Kiel
- Persing, I. & Ng, V. (2016). End-to-End Argumentation Mining in Student Essays. In Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL'16) (pp. 1384–1394).
- Quintana, C., Reiser, B., Davis, E., Krajcik, J., Fretz, E., Duncan, R., Kyza, E., Edelson, D. & Soloway, E. (2004). A Scaffolding Design Framework for Software to Support Science Inquiry. *The Journal of the Learning Sciences*, 13(3), 337–386.

- Roll, I. & Wylie, R. (2016). Evolution and Revolution in Artificial Intelligence in Education. *International Journal of Artificial Intelligence in Education*, 26(2), 582–599.
- Schiaffino, S., Garcia, P. & Amandi, A. (2008). eTeacher: Providing Personalized Assistance to E-Learning Students. *Computers and Education*, 51(4), 1744–1754.
- Somasundaran, S., Burstein, J. & Chodorow, M. (2014). Lexical Chaining for Measuring Discourse Coherence Quality in Test-taker Essays. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING'14)* (pp. 950–961).
- Stab, C. & Gurevych, I. (2017). Parsing Argumentation Structures in Persuasive Essays. *Computational Linguistics*, erscheint.
- Stegmann, K. & Fischer, F. (2016). Kurzbericht „Auswirkungen digitaler Medien auf den Wissens- und Kompetenzerwerb an der Hochschule“ (Beitrag im Rahmen eines Expertenhearings des Wissenschaftsrates).
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
- The Joint Task Force on Computing Curricula. (2013). *Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science*. Association for Computing Machinery (ACM) and IEEE Computer Society
- Wecker, C., Fischer, F. & Stegmann, K. (2012). Lern- und Kooperationsprozesse: warum sind sie interessant und wie können sie analysiert werden?. *REPORT-Zeitschrift für Weiterbildungsforschung*, 35(3), 30–41.
- Ziv, A., Wolpe, P. R., Small, S. D. & Glick, S. (2003). Simulation-Based Medical Education: An Ethical Imperative. *Academic Medicine*, 78(8), 783–787.

## Autoren



Dr. Claudia Schulz

E-Mail: [schulz@ukp.informatik.tu-darmstadt.de](mailto:schulz@ukp.informatik.tu-darmstadt.de)

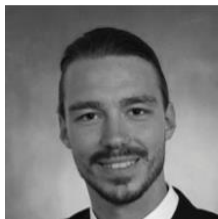
Dr. Claudia Schulz ist Postdoctoral Researcher am Ubiquitous Knowledge Processing (UKP) Lab der Technischen Universität Darmstadt. Ihre Forschungsinteressen liegen in vielfältigen Methoden aus der Künstliche Intelligenz und deren Anwendung zur Unterstützung von Entscheidungsfindung, Argumentation, und schlussfolgerndem Denken in Alltag und Bildung. Sowohl für ihre Forschung als auch ihre Lehre sowie ihr akademisches Engagement wurde sie mit Preisen ausgezeichnet. Zudem hat sie als Spezialistin für Künstliche Intelligenz bei „Outsmart Insights“ gearbeitet.



Dr. Michael Sailer

E-Mail: [michael.sailer@psy.lmu.de](mailto:michael.sailer@psy.lmu.de)

Dr. Michael Sailer ist wissenschaftlicher Mitarbeiter am Lehrstuhl Empirische Pädagogik und Pädagogische Psychologie der LMU München. Er studierte Pädagogik, Psychologie und Rechtswissenschaften mit dem Abschluss Magister Artium an der LMU München und an der University of Twente in den Niederlanden. Er promovierte in Psychologie über das Thema "Die Wirkung von Gamification auf Motivation und Leistung". Seine Forschungsschwerpunkte liegen auf der Gestaltung von simulationsbasierten Lernumgebungen, spielbasierten Lehr-Lernansätzen mit besonderem Fokus auf Gamification und Lernen mit digitalen Medien.



Dr. Jan Kiesewetter

E-Mail: [jan.kiesewetter@med.uni-muenchen.de](mailto:jan.kiesewetter@med.uni-muenchen.de)

Dr. Jan Kiesewetter ist Psychologe und Postdoctoral Researcher und Forschungskoordinator am Institut für Didaktik und Ausbildungsforschung in der Medizin am Klinikum der LMU München. Seine Forschungsinteressen liegen in individueller und kooperativer klinischer Entscheidungsfindung, Führungskompetenz in der Medizin und Argumentation. Weiterhin ist er Vorsitzender des Ausschusses für Patientensicherheit und Fehlermanagement in der Gesellschaft für medizinische Ausbildung. Sowohl für seine Forschung als auch für seine Lehrtätigkeiten wurde er mit Preisen ausgezeichnet.



Dr. Christian M. Meyer

E-Mail: [meyer@ukp.informatik.tu-darmstadt.de](mailto:meyer@ukp.informatik.tu-darmstadt.de)

Dr. Christian M. Meyer ist Senior Researcher am Ubiquitous Knowledge Processing (UKP) Lab der Technischen Universität Darmstadt. Seine Forschungsschwerpunkte liegen auf interaktiven Lernverfahren und lexikalisch-semantischen Wissensdatenbanken sowie deren Einsatz in Bildung, Journalismus und den digitalen Kultur- und Geisteswissenschaften. Er leitet den Forschungsbereich D2 im DFG-

Graduiertenkolleg „Adaptive Informationsaufbereitung aus heterogenen Quellen“ (AIPHES) und betreut eine Promotion zu intelligenten Informationssystemen. Außerdem koordiniert er das Forschungsdaten- und Wissenstransferprogramm des Kollegs und den Forschungsbereich „Writing Assistance and Language Learning“ am UKP Lab.



Prof. Dr. Iryna Gurevych

E-Mail: [gurevych@ukp.informatik.tu-darmstadt.de](mailto:gurevych@ukp.informatik.tu-darmstadt.de)

Prof. Dr. Iryna Gurevych ist Professorin für „Ubiquitäre Wissensverarbeitung“ (UKP Lab) im Fachbereich Informatik an der Technischen Universität Darmstadt. Sie betreibt Forschung im Bereich Natural Language Processing mit ihren vielfältigen Anwendungen in den Geistes-, Human- und Sozialwissenschaften. Sie ist Gründerin und Sprecherin des DFG-Graduiertenkollegs „Adaptive Informationsaufbereitung aus heterogenen Quellen“ (AIPHES) und Ko-Sprecherin des „Centrums für Digitale Forschung in den Geistes-, Sozial- und Bildungswissenschaften“ (CEDIFOR) in Darmstadt und Frankfurt.



Prof. Dr. Frank Fischer

E-Mail: [frank.fischer@psy.lmu.de](mailto:frank.fischer@psy.lmu.de)

Prof. Dr. Frank Fischer leitet den Lehrstuhl für Pädagogische Psychologie und Empirische Pädagogik an der LMU München. Er lehrt dort die Psychologie des Lehrens und Lernens in den Studiengängen der Psychologie, der Bildungswissenschaft/Pädagogik und des Lehramts. In der Forschung befasst er sich mit dem wissenschaftlichen Denken und Argumentieren, dem kollaborativen Lernen sowie mit dem Lernen mit digitalen Medien. Er war Direktor des Departments Psychologie und Dekan der Fakultät für Psychologie und Pädagogik der LMU. Seit 2009 ist er Sprecher des Munich Center of the Learning Sciences, an dem sich über 50 Wissenschaftlerinnen und Wissenschaftler aus drei Hochschulen und 8 Fakultäten in interdisziplinären Projekten in der Lernforschung und in internationalen Master- und PhD-Studiengängen engagieren.



Prof. Dr. Martin Fischer

E-Mail: [martin.fischer@med.uni-muenchen.de](mailto:martin.fischer@med.uni-muenchen.de)

Prof. Dr. Martin Fischer, MME ist Arzt und Master of Medical Education. Er ist der Direktor des Instituts für Didaktik und Ausbildungsforschung in der Medizin am Klinikum der LMU München. Er betreibt Forschung insbesondere in den Bereichen klinische Entscheidungsfindung, Wissenschaftskompetenz in der Medizin und interprofessionelle Ausbildung in der Medizin an der Schnittstelle zwischen Medizin, Pädagogik, Psychologie und Informationstechnologie. Er ist u.a. Projektleiter des Projekts FAMULUS – Förderung von Diagnosekompetenzen durch adaptive Online-Fallsimulationen in Medizin- und Lehramtsstudium, stellv. Sprecher und Projektleiter der DFG Forschergruppe COSIMA - Förderung von Diagnosekompetenzen in simulationsbasierten Lernumgebungen in der Hochschule. Zudem ist er seit 2011 der Präsident der Gesellschaft für Medizinische Ausbildung und Ars-legendi Preisträger 2015.

Das FAMULUS-Projekt ist vom Bundesministerium für Bildung und Forschung (BMBF) unter der Förderlinie „Forschung zur digitalen Hochschulbildung“ gefördert (Geschäftszeichen 16DHL1041).