

## Ausschreibung Forschungsschwerpunkt 2024

### „Mensch und generative Künstliche Intelligenz: Trust in Co-Creation“

**Hinweis:** Bitte beachten Sie unsere parallele [Ausschreibung zu Konsortialprojekten](#). Das gleiche Projekt darf nicht bei beiden Ausschreibungen eingereicht werden, aber Antragsteller:innen dürfen sich parallel mit unterschiedlichen Projekten in beiden Förderformaten bewerben.

### Hintergrund und Ansatz

Das bidt trägt dazu bei, die Entwicklungen und Herausforderungen des digitalen Wandels besser zu verstehen. Damit liefert es die Grundlagen, um die Zukunft mit der Gesellschaft verantwortungsvoll und gemeinwohlorientiert zu gestalten. Es fördert herausragende interdisziplinäre Forschung und liefert als Think Tank Entscheidungsträger:innen in Politik und Wirtschaft evidenzbasierte Empfehlungen. Das Institut fördert damit auch den offenen Dialog zwischen Forschung und Gesellschaft.

Im Jahr 2024 richtet das bidt einen neuen Forschungsschwerpunkt ein. Dieser umfasst am bidt angesiedelte interne Projekte, welche von bidt-Direktoriumsmitgliedern betreut werden, sowie externe, vom bidt geförderte Stellen (externe Projekte), die an Hochschulen und Forschungseinrichtungen in Bayern angesiedelt und dort betreut werden. In dem Forschungsschwerpunkt sollen unterschiedliche Disziplinen vertreten sein (interdisziplinärer Ansatz).

Mit der vorliegenden Ausschreibung bittet das bidt um Anträge für die externen Projekte für den Forschungsschwerpunkt. Die Förderdauer beträgt bis zu vier Jahre. Dafür stellt das bidt auf Antrag Projektmittel für die beteiligten Institutionen zur Verfügung, die bis zu 75 % der Kosten einer Stelle für Vollzeitmitarbeitende ohne Promotion (Predoc) abdecken können, für promovierte Mitarbeitende (Postdoc) bis zu 100 %.

Die Mitarbeitenden in den externen Projekten arbeiten am Standort ihrer Institution, tauschen sich aber in regelmäßigen Abständen im Forschungsschwerpunkt am bidt über ihre Forschungsaktivitäten und Ergebnisse und das weitere Vorgehen aus (agiler Ansatz).

### Forschungsschwerpunkt

Der Forschungsschwerpunkt befasst sich mit dem Thema „Mensch und generative KI: Trust in Co-Creation“.

Generative Künstliche Intelligenz (KI) bezeichnet technische Systeme, die in der Lage sind, als Reaktion auf bestimmte Eingaben Texte, Bilder, Daten oder audiovisuelle Botschaften autonom zu erstellen. Generative KI ist von großer Bedeutung, weil sie das Potential hat, zahlreiche Bereiche unseres Lebens zu transformieren, von Kunst, Design und Software-Engineering über die Transformation von Arbeitsprozessen und Berufsbildern bis hin zu Medizin, Bildung und Wissenschaft. Auch für die Gestaltung von Freizeit, politischer Beteiligung und ehrenamtlichen Engagements können fundamentale Veränderungen durch permanent verfügbare generative KI erwartet werden. Gleichzeitig stellt uns der Umgang mit generativer KI und ihre Auswirkungen auf die Gesellschaft vor vielfältige Herausforderungen, beispielsweise im Hinblick auf die Qualität und Interpretation der entstandenen Artefakte, ethische Fragestellungen, Sicherheit und potentielle soziale Ungleichheiten in Erstellung und Wahrnehmung der durch KI generierten Produkte.

Ausgangspunkt ist die Annahme, dass generative KI in den meisten Fällen als Assistenzsystem verwendet werden wird. Mensch und Technik haben dabei mindestens drei Berührungspunkte: (1) wenn Menschen mit der generativen KI interagieren, um Texte, Bilder oder audiovisuelle Botschaften zu produzieren, und/oder (2) wenn sie mit der generativen KI interagieren und nicht das Produkt der Interaktion, sondern die Interaktion selbst Wert schafft, etwa bei Verwendung generativer KI in der Lehre, der Therapie oder als Workflow-System, und/oder (3) wenn Menschen mit den von der erzeugten KI produzierten Produkten in Berührung kommen – diese beispielsweise rezipieren und/oder indem sie von der KI erzeugte Produkte in verschiedenen Kontexten verwenden und einsetzen. Der erste Punkt umfasst den schöpferischen Prozess, der gemeinsam von Mensch und Maschine durchgeführt wird (die eigentliche Ko-Kreation). Der zweite Aspekt sieht den schöpferischen Akt in der Interaktion selbst, nicht in einem dabei entstandenen Produkt. Der dritte Aspekt geht darüber hinaus und zielt auch auf den (gesellschaftlichen) Umgang mit Produkten, die gemeinsam von generativer KI und Nutzer:in erzeugt wurden.

Als zentral sehen wir den Begriff des Vertrauens an, den wir im Titel bewusst doppeldeutig formulieren. Diejenigen, die gemeinsam mit der Maschine Inhalte erzeugen (die Produzent:innen), wollen der Arbeitsweise und den Vorschlägen der generativen KI vertrauen können: trust in co-creation verstanden als „Vertrauen in Prozesse und Artefakte während der Ko-Kreation“. Die Rezipient:innen der Produkte möchten den erzeugten Inhalten vertrauen können und Vertrauen ist zudem eine wichtige Vorbedingung dafür, dass die erzeugten Produkte in verschiedenen Kontexten eingesetzt werden: trust in co-creation verstanden als „Vertrauen in die Produkte der Ko-Kreation“.

Vertrauen ist ein wichtiger Bestandteil von Beziehungen und Interaktionen mit anderen. Ohne Vertrauen würden wir uns in ständiger Sorge vor Betrug und Verletzungen isolieren. Jemandem zu vertrauen, setzt dabei voraus, dass wir ihn oder sie als vertrauenswürdig einschätzen. Dabei ist jedoch zwischen tatsächlicher und wahrgenommener Vertrauenswürdigkeit zu unterscheiden: Man kann jemandem vertrauen, obwohl dieser nicht vertrauenswürdig ist („misplaced trust“) bzw. nicht vertrauen, obwohl es eigentlich eine vertrauenswürdige Person wäre („misplaced mistrust“). Diese Punkte lassen sich auch auf technische Systeme übertragen: Vertrauen in ein technisches System basiert einerseits auf Voreinstellungen, die durch generelles Vertrauen in technische Systeme begründet sind und andererseits auf Erfahrungen, die in der Interaktion mit einem System gemacht werden. Vertrauenswürdigkeit von KI-Systemen ist die zentrale Leitlinie im Europäischen AI Act. Hier werden sieben Bewertungskriterien angeführt: (1) Vorrang menschlichen Handelns und menschliche Aufsicht, (2) Technische Robustheit, Genauigkeit und Sicherheit, (3) Datenschutz und Datenqualitätsmanagement, (4) Transparenz, Nachverfolgbarkeit und Erklärbarkeit, (5) Vielfalt, Nichtdiskriminierung und Fairness, (6) Gesellschaftliches und ökologisches Wohlergehen, Nachhaltigkeit, Gesellschaft und Demokratie, und (7) Rechenschaftspflicht und Nachprüfbarkeit. Neben der faktischen Vertrauenswürdigkeit spielt auch die wahrgenommene Vertrauenswürdigkeit eines Systems eine Rolle. Als Einflussgrößen auf letztere werden insbesondere folgende genannt: Die Einschätzung der Leistungsfähigkeit eines Systems u.a. bzgl. Korrektheit, die Nachvollziehbarkeit und Vorhersehbarkeit von Systemverhalten, sowie die vermuteten Absichten, die die Entwickler:innen mit dem System verfolgen.

## Dimensionen des Forschungsschwerpunkts

Das Forschungsprogramm ist entlang der folgenden Dimensionen strukturiert:

1. Prozess der Ko-Kreation (produzieren, interagieren und rezipieren)
2. Grundsatzbereich
3. Anwendungsfeld

Forschungsprojekte müssen sich durch die Formulierung konkreter, gesellschaftlich relevanter Forschungsfragen in jeder der Dimensionen positionieren.

Wie oben beschrieben, spielt „Vertrauen“ in allen drei Ausprägungen des **Prozesses der Ko-Kreation** eine Rolle: In der Interaktion mit der generativen KI, wenn die Interaktion als solche Wert stiftet; in der Produktion und Anwendung der von der KI generierten Texte, Bilder oder audiovisueller Inhalte; und in der Rezeption der von generativer KI erzeugten Produkte. Dabei beeinflussen frühere Interaktionen und Erfahrungen nachfolgende Interaktions- und Rezeptionssituationen.

Den Kern des Forschungsprogramms bilden folgende **Grundsatzbereiche**:

1. Bedingungen für Vertrauen: In diesem Grundsatzbereich geht es darum, unter welchen Bedingungen der KI selbst oder den erzeugten Produkten vertraut wird. Bedingungen können dabei beispielsweise technisch-gestalterischer Natur sein (z. B. bezogen auf die verwendeten Modelle oder Interfaces), können sich auf den/die Nutzer:in selbst beziehen (z. B. Einfluss von Vorerfahrungen, Kompetenzen), das erzeugte Produkt betreffen (bspw. Aussehen, Qualität der Inhalte, Transparenz, Verwendung von Quellen) oder dessen Anwendung (relevant für Vertrauensbedingungen sind hier bspw. beruflicher/gesellschaftlicher Kontext, Umfeld/Team). Ebenso relevant sind Grenzen des Vertrauens und (Gestaltungs-)Möglichkeiten, um das Vertrauen in generative KI zu stärken.
2. (Wahrgenommene) Güte/Qualität; Gütemaße/Qualitätskriterien: Dieser Grundsatzbereich umfasst alle Aspekte der Güte oder Qualität. Dies spielt eine Rolle bei der Interaktion (z. B. Qualität der während der Interaktion zur Verfügung gestellten Informationen, bspw. die Korrektheit von Lerninhalten oder die Adäquatheit von vorgeschlagenen Schritten im Workflow), bei der Beurteilung des erzeugten Produkts und natürlich bei der Rezeption der von der generativen KI erzeugten Produkte. Relevante Fragestellungen dabei sind: Was bedeutet „gut“ überhaupt, wie kann man das messen, was ist „gut genug“, und in welchen Kontexten ist das wichtig? Was sind Akzeptanzkriterien beim Schaffen und beim Konsumieren der Inhalte? Wie legen wir Schwellwerte für „gut genug“ fest? Wie stellen wir sicher, dass Informationen, die von generativer KI erzeugt werden, richtig sind?
3. (Erforderliche) Kompetenzen/Literacy: Sowohl für die vertrauensvolle Interaktion und Produktion, als auch für die Rezeption braucht es unterschiedliche Kompetenzen auf Seiten der Nutzer:innen. Wie und durch wen kann die Güte von Texten, Bildern, Code, Videos, im beruflichen, privaten, politischen, ausbildungsbezogenen Kontext überprüft werden, welche Fähigkeiten und Fertigkeiten benötigen wir dafür, wie können wir diese vermitteln oder trainieren? Welche Hürden und Ungleichheiten zeigen sich hinsichtlich der Kompetenzen zwischen verschiedenen Bevölkerungsgruppen?
4. Veränderung von Rollen, Identitäten, Aufgaben: Der vertrauensvolle Umgang mit und der Einsatz von generativer KI in unterschiedlichen Lebensbereichen wird eine Veränderung von Rollen, Identitäten und Aufgaben mit sich bringen. Dies ist besonders im beruflichen Kontext relevant – so stellt sich beispielsweise die Frage, wie sich Berufsbilder und -rollen verändern werden.
5. Technischer, ethischer und normativer Rahmen: Dieser Grundsatzbereich zielt darauf ab, welche technischen, ethischen und normativen Rahmenbedingungen nötig sind, um einen vertrauensvollen Umgang mit generativer KI zu gewährleisten. Dies kann bspw. Fragen danach umfassen, wie Sicherheitsmechanismen eingebaut werden können, um Missbrauch oder Fehlinterpretation von generativen KI-Modellen zu verhindern oder auch an welcher Stelle bewusst Grenzen für den Einsatz generativer KI definiert werden sollten – sollen bspw. Demenzkranke mit ChatGPT interagieren dürfen (Menschenwürde vs. ggf. therapeutische Effekte)? Darüber hinaus stellen sich im Kontext des Einsatzes von generativer KI viele rechtliche Fragen, beispielsweise im Umfeld des Urheberrechts oder des Persönlichkeitsschutzes, aber auch im Prüfungsrecht.

Die Grundsatzgebiete definieren abstrakte Forschungsfragen, sind aber zu allgemein, um zielgerichtet beantwortet werden zu können. Sie sollen deswegen in verschiedenen **Anwendungsfeldern** konkretisiert

werden. Diese Anwendungsfelder sollen in den drei am bidt bestehenden Forschungsbereichen verortet werden:

- Staat, Regulierung und Infrastruktur
- Kommunikation, Gesellschaft und Partizipation
- Wirtschaft und Arbeitswelt

Zusammenfassend muss sich folglich jedes Projekt mit einer der drei konkreten Fragestellungen im Prozess der Ko-Kreation (Interaktion, Produktion oder Rezeption) befassen, einen der fünf Grundsatzbereiche adressieren und sich auf ein konkretes Anwendungsfeld innerhalb der drei Forschungsbereiche beziehen.

Hierzu bittet das bidt um die Einreichung von Anträgen für externe Projekte.

## Förderformat

Ein typisches Projekt läuft über vier Jahre und umfasst eine:n Antragsteller:in und eine:n wissenschaftliche:n Mitarbeiter:in. In besonders begründeten Ausnahmefällen kommen auch Projekte mit zwei Antragsteller:innen mit je einem Mitarbeitenden in Betracht. In diesem Falle müssen die Antragsteller:innen unterschiedlichen wissenschaftlichen Disziplinen angehören.

Antragsberechtigt sind Wissenschaftler:innen, die eine abgeschlossene Promotion nachweisen können. Die Antragsteller:innen sowie deren über das Projekt geförderte Mitarbeitende müssen eine Anbindung an eine wissenschaftliche Einrichtung mit Sitz in Bayern aufweisen (in der Regel Universität, Hochschule oder außeruniversitäre Forschungseinrichtung), die spätestens zum geplanten Projektstart besteht.

Finanziert werden die Kosten für eine:n wissenschaftliche:n Mitarbeiter:in der Antragsteller:innen. Dabei beträgt der zuwendungsfähige Beschäftigungsumfang der wissenschaftlichen Mitarbeiter:innen bei Promovierten max. 100 %, bei Nichtpromovierten max. 75 % eines Vollzeitäquivalents. Die Stellen der Antragstellenden selbst sind nicht förderfähig. Doppelförderungen sind ausgeschlossen. Werden nichtpromovierte Mitarbeiter:innen in Teilzeit angestellt, übernimmt das bidt somit die kompletten Personalkosten, allerdings nur bis zu einer Höhe von 75 % einer Vollzeitstelle. Bei einer Vollzeitbeschäftigung von nicht promovierten Mitarbeiterinnen und Mitarbeitern müssen die restlichen Mittel durch die Antragsteller:innen aufgebracht werden. Die wissenschaftlichen Mitarbeiter:innen werden an der Institution der Antragsteller:innen angestellt.

Die Antragsteller:innen sowie deren Mitarbeitende nehmen regelmäßig an Treffen des Forschungsschwerpunkts teil. Einmal im Halbjahr finden eineinhalbtägige Workshops in Präsenz (in der Regel in München) statt, an welchen Antragstellende und Mitarbeitende teilnehmen. Zeitlich versetzt findet einmal im Halbjahr ein eintägiges Treffen der Mitarbeitenden in Präsenz (in der Regel in München) statt. In Monaten ohne solche Treffen tauschen sich die Mitarbeitenden in einem virtuelle Jour Fixe aus.

In Ansatz gebracht werden können Personalkosten für die Stelle und zudem Sachmittel (z. B. für Reisen, Veranstaltungen und wissenschaftliche Hilfskräfte) sowie eine Kostenpauschale für die Nutzung der Infrastruktur an der Institution der Antragsteller:innen (Overhead). Im Projektantrag können die Sachmittel in Höhe von max. 10.000 Euro pro Antragsteller:in und Jahr, die Infrastrukturpauschale in Höhe von maximal 20 % der Personalkosten angesetzt werden. Nur in besonders begründeten Ausnahmefällen können durch die Projektspezifika bedingte zusätzliche Sachkosten für das Gesamtprojekt beantragt werden (bspw. Umfragen mit sehr hohen Teilnehmendenzahlen in mehreren Ländern, besondere IT-Ausstattung). Zahlungen erfolgen nur für tatsächlich im Projektzeitraum angefallene Kosten, die Infrastrukturpauschale wird in Höhe von maximal 20 % der tatsächlichen Personalkosten geleistet. Aus dem Sachmittelbudget sind anfallende Reisekosten für die Präsenztermine des Forschungsschwerpunkts zu finanzieren.

# Auswahlverfahren und Bewilligung

Das Auswahlverfahren erfolgt zweistufig: In der ersten Stufe sind eine Projektskizze sowie Lebensläufe einzureichen. Das bidt-Direktorium entscheidet daraufhin, welche Einreichungen für die zweite Stufe zugelassen werden und einen Vollertrag einreichen können.

## Erste Stufe: Projektskizze

Die Projektskizze für ein Projekt soll vier Seiten – in deutscher oder englischer Sprache – umfassen und folgende Aspekte abdecken:

- Thema
- Antragsteller:in inkl. Institution
- Kontext und Problem inkl. Bezug zu den in der Ausschreibung aufgezeigten Dimensionen
- Lösungsideen, Lösungsansatz und erwarteter Output
- Erwarteter wissenschaftlicher und gesellschaftlicher Beitrag
- Einbindung der Öffentlichkeit und von Stakeholdergruppen zur Förderung des Austauschs von Forschung und Gesellschaft
- Stellenbesetzung (Zeitraum, Stellenanteil Vollzeitäquivalent, Predoc/Postdoc-Rolle und Gehaltsgruppe/-stufe der Stelle)

Zusätzlich ist der Lebenslauf der Antragsteller:in einzureichen (eine Seite).

## Zweite Stufe: Vollertrag

Der Vollertrag für ein Projekt soll zehn Seiten – in deutscher oder englischer Sprache – umfassen und, zusätzlich zu den oben für die Projektskizze genannten Punkten, folgende Aspekte abdecken:

- Detaillierter Projekt- und Arbeitsplan, inklusive wichtiger Meilensteine des Projekts (insb. Meilensteine zur Evaluation nach der ersten Hälfte der Projektlaufzeit)
- Detaillierter Kostenplan (gemäß bereitgestellter Vorlage zur Kostenkalkulation): Kosten für Personal, Sachmittel und Infrastruktur für die an den beteiligten Institutionen anzustellenden Projektmitarbeitenden (pauschalisierte Ansätze, Personalkosten nach den aktuellen Durchschnittskostensätzen)
- Bestätigung, dass – sofern für das Vorhaben nach den Richtlinien Ihrer Institutionen erforderlich – Ethikvoten eingeholt wurden und diese der Durchführung des Projekts nicht im Wege stehen. Etwaig erforderliche Voten müssen spätestens zum geplanten Projektbeginn vorliegen.

## Auswahlkriterien

Die Einreichungen zur ersten und zur zweiten Stufe werden anhand folgender Kriterien bewertet:

- Ausmaß des zu erwartenden wissenschaftlichen Beitrags
- Gesellschaftliche und politische Relevanz der zu erwartenden Ergebnisse
- Bezug zu den für diese Ausschreibung definierten Forschungsbereichen
- Einbindung Öffentlichkeit/Stakeholdergruppen
- Wissenschaftlicher Ausweis der Antragstellenden
- Umsetzbarkeit des Projekts

Die Entscheidung über die Zulassung zur zweiten Stufe (Einreichung eines Vollertrags) sowie über die Bewilligung der Konsortialprojekte in der zweiten Stufe trifft das bidt-Direktorium anhand der aufgeführten

Kriterien. Für die Entscheidung in der zweiten Stufe werden externe wissenschaftliche Gutachten hinzugezogen.

## Bewilligung

Akzeptierte externe Projekte erhalten zunächst eine Bewilligung für eine Förderung über zwei Jahre. Eine Förderung im dritten und vierten Jahr erfolgt, wenn die nach 1,5 Jahren vom bidt durchzuführende Evaluation positiv verläuft. Dazu sind die nach 1,5 Jahren der Projektlaufzeit zu erwartenden Ergebnisse als Meilensteine präzise zu benennen.

Die Zuwendung erfolgt auf Basis der Allgemeinen Nebenbestimmungen für Zuwendungen zur Projektförderung (ANBest-P). Die Bewilligung der externen Projekte durch die Bayerische Akademie der Wissenschaften erfolgt gegenüber der Institution der Antragsteller:in.

## Zeitplan

7. Mai 2024 15:00 Uhr MESZ	Einreichungsfrist für die erste Stufe (Bewerbungsformular, Projektskizze mit Lebenslauf)
14. Juni 2024	Mitteilung über die Zulassung zur zweiten Stufe
9. September 2024 15:00 Uhr MESZ	Einreichungsfrist für die zweite Stufe (Vollantrag)
Mitte November 2024	Mitteilung über die Bewilligung
Erstes Quartal 2025	Projektstart

## Kontakt

Informationen zur Ausschreibung, zur Einreichung sowie eine Sammlung häufiger Fragen finden Sie unter: <https://www.bidt.digital/foerderprogramm/forschungsschwerpunkt-generative-ki/>

Sie finden die Datenschutzerklärung unter <https://www.bidt.digital/datenschutzerklaerung/>

Bei Fragen zur Ausschreibung wenden Sie sich gerne an: [forschung@bidt.digital](mailto:forschung@bidt.digital)