

⇒ Sylvia Kühne

Playful technologies? Über Anspruch und Risiko von Künstlicher Intelligenz im Wargaming

⇒ Schlagworte

Wargame, Epistemologie, LLMs, Künstliche Intelligenz, Ethik

⇒ 1 Einleitung:

»Mass Casualty Event«
Watching from inside
the game, billowing gasses
for camouflage, clouds
of green, luminous clouds
of yellow, the flushed face
of the Major's child, our faces,
white-gold torches in the meadow.

*I am in War. No,
I am in a game
Of war. No, I am in a painting.*

Nomi Stone (2019, 54, Herv. d. Verf.)

Sylvia Kühne, Dr., Studium der Soziologie und Kriminologie, Wissenschaftliche Mitarbeiterin im AI Ethics Research Hub, Lehrstuhl für Sozial- und Technikethik an der Helmut-Schmidt-Universität/Universität der Bundeswehr Hamburg. Neuere Veröffentlichung: Gut Feelings and Algorithms. Searching for Harmful Intentions in Airport Security Processes. In: Engaging Science, Technology, and Society 10(3) 2024 (zus. mit Bettina Paul): DOI: 10.17351/ests2023.2337

ORCID: 0009-0002-2992-4920

DOI: [10.18156/eug-1-2026-art-6](https://doi.org/10.18156/eug-1-2026-art-6)

Aus den wirklichen Unwirklichkeiten des fiktiven Landes »Pine-land« heraus erzählt Nomi Stone in ihrem Gedichtband »Kill Class« (2019) mit der Stimme der Anthropologin davon, was es bedeutet die Schrecken des Krieges nachzuspielen. Im Rahmen ihrer zweijährigen ethnografischen Feldforschung in diesem ländlichen, nah-

östlich anmutenden Raum in North Carolina, in dem das US-Militär Vorbereitungsübungen für den Irak-Einsatz durchführt, begleitete sie Soldat:innen und für die Szenarien angeheuerte, aus dem Nahen Osten stammende Zivilist:innen, die als Stellvertreter:innen für die Bevölkerung des Einsatzgebietes fungierten. Im Verlauf von Stones Gedichten zeigt sich, wie diese Stimme gleichermaßen durch (An-)Teilnahme¹ immer mehr in das Wargame hineingezogen wird, in ihrer Positionierung aber immer auch außen vor verbleibt. Was einerseits als Zurückweichen vor dem Erlernen von Gewalt gelesen werden kann, tritt andererseits auch als ein Problem zutage, das das ethnografische Prinzip des »going native« nicht nur reflektiert, sondern darüber hinausgeht: Ihre Rolle der Beobachterin gefährdet die Existenz der Realität des Spiels, während umgekehrt das Mitspielen die analytische Aufmerksamkeit für das Geschehen verhindert. Dieses Paradox von Immersion und Abstraktion, das sich in dieser Form des Wargaming entfaltet, illustriert Stone ausdrucksvoll im eingangs zitierten Gedicht »Mass Casualty Event«. Insbesondere die drei letzten Zeilen thematisieren, wie ihr anthropologisches Augenmerk von der Theatralisierung eingenommen ist, sodass scheinbar verschwimmt, was Wirklichkeit, was Spiel ist und sodann die Frage eröffnet, was es ist, von dem sie vereinnahmt wird? Im – insofern vielleicht auch als Rüstzeug angelegten² – poetischen Zugang zum Phänomen Wargame illustriert sich kunstvoll und reflexiv, wie hier Realität und Spiel ineinander verschränkt sind und dadurch neue Erfahrungswelten erschaffen werden.

Jean Baudrillard (2010) hat hierfür den Begriff der Hyperrealität geprägt, der darauf verweist, wie Simulationen oder Bilder der Realität die tatsächliche Realität verdrängen oder sogar ununterscheidbar von ihr werden lassen können. Ihm (1991) zufolge erzeugen Wargames – seien es aktive spielerische Exerzitien in einer inszenierten,

(1) Diese Anteilnahme bezieht sich vor allem auf die zivilen Rollenspieler:innen, die mit dem Anspruch der Trainingssimulation konfrontiert werden, authentisches kulturelles Wissen zu verkörpern und gleichzeitig gehalten sind, sich den Vorgaben der militärischen Rollenspiele zu fügen. Hierin erkennt Stone (Hagman/Sacco 2017) eine koloniale Logik kultureller Übersetzung, in der das tatsächliche Wissen des/der Anderen nicht relevant ist.

(2) Der Begriff des Rüstzeugs ist hier in Anlehnung an Loïc Wacquants Hinweis (2009, 119, H.i.O.) an ethnografische Forscher:innen – »go native but ›go native armed« – gebraucht. Stone selbst hat ihrem 2019 erschienen Gedichtband 2022 mit »Pinelandia. An Anthropology and Field Poetics of War and Empire« eine anthropologische Studie zu ihren Feldbeobachtungen folgen lassen, in der sie nicht nur ausführlich die Widersprüche in den Erwartungen an die Authentizität von Trainingssimulationen thematisiert. Sie ordnet hierin auch die »field poetry« als Zugangsweise zur Situiertheit der Anthropologin als ein »enacted genre« (212) ein.

und wie bei Stone nachgebauten, Umwelt oder klassische Wargames in Form des Brettspiels und seiner digitalen Äquivalente, wie sie Gegenstand dieses Beitrages sind –, eine Hyperrealität des Krieges, in der dieser als Simulation realer wird als reale Konflikte: ein Simulacrum, das sich aus der Simulation militärischer Übungen speist (vgl. Hirst 2024). Wargames sind aber nicht nur die Mittel, mit denen sich ein Phänomen wie Krieg realisiert oder mit Hilfe derer Militär seine Aktionen plant. Mit ihnen sollen Strategie und Doktrin für seine Spieler:innen zudem konkret erfahrbar werden. Wie im Gedicht von Stone entfaltet, gilt vor diesem Hintergrund ein spezifisches Verhältnis von synthetischer und immersiver Erfahrung als zentrales Merkmal des Wargaming (vgl. ausführlich Hirst 2024), wie es sich auch im Begriff des »experimental realism« (McDermott 2002, 332–334) niederschlägt. Mit ihm ist die Verschmelzung von Abstraktheit – hier der militärischen Kriegssituation, exemplifiziert durch ein kontextuelles Spielszenario mit Möglichkeiten von Aktion und Gegenaktion im Rahmen szenariobasierter und mal mehr, mal weniger rigider Spielregeln (»Game«) –, und Konkretheit als damit einhergehendem Erfahrungsraum der Spieler:innen, die auf diese Weise die eigenen Handlungen ohne physischen Schaden sowohl kognitiv als auch emotional erleben können sollen (»Play«), beschrieben.

Gegenwärtig ist die Ermöglichung dieses Erfahrungsraumes aktueller denn je, denn »Wargaming is back in vogue« (Barzashka 2019). Wie insbesondere Aggie Hirst, die 2024 mit ihrer dekonstruktiven Analyse das jüngst wohl umfassendste Werk zum Wargaming vorlegte, gezeigt hat, bildet dieses spezifische Verhältnis von Game und Play, einer im deutschen Sprachgebrauch nicht üblichen Trennung der Bedeutungsebenen des Begriffes Spiel, einen argumentativen Kern für die neuerlichen Renaissance dieses längst historischen »Managementwerkzeugs« (Nitzl et al. 2023, 348) im Militär. Vor dem Hintergrund der KI-Revolution und einem Sicherheitsumfeld, das als von weit verbreiteter Mehrdeutigkeit und Komplexität geprägt charakterisiert wird, soll es dazu dienen, »the messy business of human thinking and decision-making« (Hirst 2024, 79) zu adressieren. Dem Verhältnis von Game and Play im Wargame nachzuspüren, wie es das Anliegen dieses Beitrages ist, erscheint dann auch vor allem deshalb relevant, weil Technologien im Feld der Künstlichen Intelligenz (KI), nicht nur den Hintergrund für seine gegenwärtige Renaissance bilden. Sie sollen selbst Eingang in das Wargaming finden. Wenn also Befürworter:innen von Large Language Models (LLMs) im Wargaming gegenwärtig mit dem Argument antreten, diese Unordnung menschlicher Entscheidungsprozesse durch Effizienz und Genauigkeit zu beseitigen, um zuverlässige,

objektive Entscheidungsprozesse prädiktiv zu modellieren (vgl. z.B. Chen und Chu 2024, 2894), dann fokussiert dieser Beitrag, ob und wie sich mit LLMs der epistemische, im Verhältnis von Game und Play konstituierende, Charakter des Wargaming verändert und mithin sein ihm zugeschriebenes Potenzial, durch den im Spiel eröffneten Erfahrungsraum Ungewissheit erlebbar zu machen und relevante Lernprozesse anzustoßen (vgl. Hirst 2024). Weniger also an der Frage nach der Logik spielerischer Gewalt selbst, der Bedeutung digitaler Technologie als distanzierterem Werkzeug zur Trivialisierung und Gamifizierung des Todes (vgl. dazu zum Beispiel Pugliese 2016; 2017) oder der oft wiederholten Frage, ob sogenannte klassische Wargames im Zeitalter digitalisierter Kriegführung und wachsender systemischer Abhängigkeiten überhaupt noch geeignet sind, neu entstehende Formen von Komplexität und Unsicherheit zu erfassen (vgl. Weller et al. 2024), schlägt dieser Beitrag folgenden Weg ein: Er folgt den Argumenten der Verfechter:innen des Wargaming als militärischem Trainingstool und fragt nach den Konsequenzen, wenn mit der KI-Technologie als Bestandteil des Instrumentes Wargame ein neues, gleichfalls wirklichkeitshervorbringendes Element in die Konstellation von Game und Play eintritt.

Hierfür wird im ersten Abschnitt das epistemologische Prinzip des Wargaming skizziert (2), bevor sich daran eine Darstellung der Automatisierungsbestrebungen von Wargames anschließt (3), in der die sich aus der neuerlichen Renaissance des Wargames ergebenden Erwartungen an ihre technische Agency als auch die Versprechen und Anwendungsideen für LLMs in Wargaming systematisiert werden. Die Folgen für das Verhältnis von menschlicher und technischer Agency im technisierten Wargame befragend, werden dann die Implikationen letzterer diskutiert (4) und gezeigt, dass diese geeignet sind, nicht nur in epistemologischer Hinsicht den pädagogischen Anspruch des Wargamings zu unterlaufen, sondern auch ein ihm implizites moralisches Moment zu konterkarieren.

⇒ 2 Epistemologische Skizze militärischer Wargames

Unter Wargames werden im Allgemeinen Instrumente verstanden, die die menschliche Entscheidungsfindung unter Bedingungen unvollständiger oder imperfekter Informationen erlebbar werden und mithin trainieren sollen (vgl. z.B. Hirst 2024; Rubel 2006; Sabin 2014), indem sie darauf abzielen, einen spezifischen Erfahrungsraum zu konstruieren, der zugleich fiktiv und realistisch ist. Konstitutiv für diesen Raum ist die Beziehung zwischen synthetischer und immersiver Erfahrung – die, wie

im Folgenden ausgeführt, als argumentativer Kern für die Relevanz der Wargame-Praxis fungiert.

⇒ 2.1 Wargaming als synthetische Erfahrung

Seit mehr als einem Jahrhundert gilt Wargaming im militärischen Kontext als zentrales Mittel zur künstlichen Antizipation von Krieg – ein Phänomen, das durch Komplexität und Unvorhersehbarkeit des Ausgangs und insofern eine enorme Unbestimmtheit gekennzeichnet ist – sowohl in der Ausbildung als auch in der strategischen Planung (vgl. Hirst 2024; Sabin 2014; Perla 1990). Von der Verwendung topografischer Karten im preußischen Kriegsspiel (1824) über klassische Tabletop- und Seminar-Wargames bis hin zu digitalen Formaten in jüngerer Zeit wird der übergeordnete Zweck von Wargames darin gesehen, auf dem Schlachtfeld der Zukunft einen entscheidenden Vorteil zu erlangen – und zwar nicht nur durch die retrospektive Auseinandersetzung mit vergangenen Konflikten, sondern vor allem durch eine idealisierte Vorwegnahme möglicher Entscheidungsspielräume zukünftiger Kampf- und Krisenszenarien (für einen historischen Überblick über die Entwicklung von Wargames vgl. z. B. Hirst 2024, 58–65; Perla 1990). Mit anderen Worten zielt Wargaming darauf ab, das Potenzial einer Vorgehensweise sowie deren Grenzen im Lichte gegensätzlicher Fähigkeiten und Handlungen und der Ungewissheit ihres Ausgangs zu *visualisieren*, *abzuwägen* und zu *bewerten* – mithin sie auf diese Weise erfahrbar werden zu lassen (vgl. Mouat 2022, 5).

In diesem Sinne fungiert Wargaming in der militärischen Ausbildung als eine Verschmelzung von Abstraktion – des militärischen Konfliktszenarios selbst, kontextualisiert durch eine spezifische Spielumgebung, die Möglichkeiten zum Handeln und Reagieren bietet und mehr oder weniger starren Regeln unterliegt – und der entsprechenden Erfahrungsdimension, in der die Spieler:innen die Konsequenzen ihrer Entscheidungen ohne physischen Schaden erfahren (vgl. z.B. Huizinga 2022; Sicart 2014). Diese spezifische Beziehung zwischen Abstraktion und authentischer Erfahrung spiegelt sich in der terminologischen Mehrdeutigkeit von Wargames und Simulationen wider, sowohl im professionellen als auch im akademischen Diskurs (vgl. z.B. Hirst 2024, 75; Sabin 2014, 4–17.). In diesem Zusammenhang findet sie weiteren Ausdruck in den anhaltenden Debatten über die ökologische Validität der Methode Wargame, die sich auch an der Frage ausmacht, ob der Begriff »Simulation« oder jener des »Spiel« einer (Darstellung der) Wirklichkeit näherkommt (vgl. Hirst 2024, 23–29.). Für Hirst (ebd., 143–144) ergibt

sich diese Ambivalenz wesentlich aus den semantischen Assoziationen von Simulation und Spiel, insbesondere in der, im deutschen Sprachgebrauch unüblichen Differenzierung von Game und Play – Begriffe, die wiederum für das Verständnis der epistemischen und pädagogischen Ansprüche von Wargames in der militärischen Ausbildung von wesentlicher Bedeutung sind. Danach wird Simulation aufgrund ihrer Assoziation mit den Naturwissenschaften typischerweise als objektiv wahrgenommen. Etymologisch dem lateinischen *simulāre* entlehnt und auf Nachahmung, Vortäuschen, Scheinheiligkeit verweisend (DWDS 2025a) – fand der Begriff in der zweiten Hälfte des 20. Jahrhunderts Eingang in die wissenschaftliche Terminologie und bezeichnet die genaue Modellierung technischer oder natürlicher Prozesse. Heute fungieren Simulationen als wissenschaftliche Werkzeuge zur Generierung von Wissen über zukünftige Entwicklungen (Kaminski et al. 2023, 14) und sind oft eng mit Computersimulationen verknüpft, in denen komplexe, reale Systeme reduziert werden, um relevante Systemdynamiken im Zeitverlauf zu modellieren oder komplexe Systemzustände zu einem bestimmten Zeitpunkt darzustellen. Demgegenüber gilt das Konzept des Spiels als sehr viel subjektiver – im Deutschen etwa als eine »nicht auf Nutzen ausgerichtete, vergnügliche, mit Ernst betriebene Tätigkeit« oder »Zeitvertreib« (DWDS 2025b) –, eine Bedeutung, die sich analog in den etymologischen Bedeutungen von Game und Play findet: Abgeleitet vom Altenglischen *gamen*, bedeutet ersteres so viel wie »Freude, Spaß, Spiel, Vergnügen« (Harper 2025a) und erhielt um 1200 die zusätzliche Bedeutung als »Wettstreit um Erfolg oder Überlegenheit nach Regeln« (ebd.). Es ist aber insbesondere der Begriff des Play, so Hirst (2024, 59) resp. die Begrifflichkeit *Gameplay* (ebd. 139), die die oben angedeutete Ambivalenz des Wargaming begründet. Etymologisch vom Germanischen *plegōjanan*, »mit sich selbst beschäftigt sein« (Harper 2025b) stammend, indiziert der Begriff oft eine subtile Abwertung aufgrund seiner Assoziation mit Kindheit, Vergnügen oder unernster Beschäftigung.

Die Begriffe Simulation, Game und Play wiederum zueinander ins Verhältnis setzend, lässt sich argumentieren, dass Spiele – im Sinne des Games – ebenso wie Simulationen eine synthetische Umwelt kreieren und insofern als »Realitätsverdopplung« fungieren, die, Niklas Luhmann (2017, 67–80) folgend, episodisch und durch spezifische Abstraktionsbedingungen (Szenario, Regeln, zeitliche Begrenzungen) gekennzeichnet sind. Insofern beziehen sich sowohl Spiel als auch Simulation konzeptionell auf die Reproduktion eines Prozesses und auf den Versuch, Verhalten unter variierenden Bedingungen zu untersuchen und zu analysieren, ohne das reale – oder genauer gesagt das

antizipierte – System direkt beeinflussen zu können. Beide basieren auf Modellen, in denen die wesentlichen Merkmale und Zusammenhänge des »realen« (retrospektiven) oder antizipierten (prospektiven) Systems in abstrahierter Form dargestellt werden. Im Gegensatz zu Simulationen ermöglicht das Gameplay jedoch eine Erfahrung, die sich zugleich auf eine Realität bezieht als auch von ihr distanziert. Dieses Absetzen der »realen Realität« von der »fiktiven« Realität (Luhmann 2017, 68) wird in der konzeptionellen Trennung von Game und Play aufgefangen. Während der »Gamemodus« (Walther 2003) strukturell und zeitlich begrenzt ist, ist es das Play, das dem Spiel seine erfahrbare Bedeutung verleiht: »While games impose order, structure, and boundaries, play operates through and promotes creativity and agency« (Hirst 2024, 143).

⇒ 2.2 Wargaming als immersives Erleben

Vor diesem Hintergrund sind es die dem Play innewohnenden sogenannten Flow-Erlebnisse und die Immersion (vgl. hierzu ausführlich Hirst 2024, 259–284.), die das so tun *als-ob* des Games in eine Form natürlicher und notwendiger Handlung verwandeln. Diese Elemente heben die Abstraktion des Modells auf und machen die Subjektivität des Spielers/der Spielerin zum zentralen Bestandteil der Wargaming-Praxis. Dementsprechend gilt die Integration mindestens eines/einer menschlichen Entscheidungsträgers/-trägerin als entscheidender Unterschied zwischen einer Simulation und einem Wargame (ebd., 23).

Es ist das Ziel eines Wargames *menschliche* Spieler:innen, basierend auf impliziten oder expliziten Regeln, in eine kompetitive Umgebung eintauchen zu lassen, um sich mit den möglichen Konsequenzen ihrer Handlungen auseinanderzusetzen (vgl. z.B. Lin-Greenberg et al. 2022, 85). Diese Grundlegung, die so auch von Hirst (2024) aufgegriffen wird, spiegelt sich in der einflussreichen Definition von Peter Perla (1990, 164) – einer zentralen Figur sowohl in der Wargaming-Community als auch der Wargaming-Theorie – wider. Danach ist ein Wargame »an exercise in human interaction, and the interplay of human decisions and the simulated outcomes of those decisions.« Dieses Zusammenspiel menschlicher Entscheidungen und deren simuliertes Ergebnis mache es unmöglich, dass zwei Spiele einander gleichen. Simulationen hingegen eigneten sich daher besser für die Untersuchung anderer, technischerer Aspekte der Realität (ebd.). Und weiter: »[w]ithout human players there may be a model, but there is no game« (Perla 1990, 274). In diesem Zusammenhang kommen auch Erik Lin-

Greenberg et al. (2022, 85–87.) zu dem Schluss, dass das menschliche Element von Wargames sie von Computersimulationen oder ökonomischen ›Spielen‹ unterscheidet, in denen Modelle angenommenes menschliches Verhalten simulieren.

Das Zusammenspiel von Abstraktion (Game) und Immersion (Play) gilt allen Formen von Wargames inhärent, unabhängig davon, ob sie etwa als Bildungs- oder analytische Wargames klassifiziert werden – eine Unterscheidung, die, wie Hirst (2024, 55) betont, weder in der Wargaming-Community noch in der wissenschaftlichen Literatur eine distinkte Kategorisierung darstellt, schon allein deshalb, weil, wie sie vor diesem Hintergrund argumentiert, alle Spiele »einen pädagogischen Zweck haben« (ebd., 55; vgl. ebd., Kapitel 5; vgl. auch Rubel 2006). Das heißt, dieser Zweck gilt unabhängig davon, ob er im Prozess oder im Produkt des Spiels liegt – ebenso wie er über andere taxonomische Rahmen hinausgeht, wie etwa die Unterscheidung zwischen politisch-militärischen Spielen, militärisch-politischen Spielen, Militärspielen und analytischen Übungen, wie sie von Paul K. Davis und Paul Bracken (2025, 34) vorgeschlagen wird. In der Praxis manifestiert sich diese konzeptionelle Konvergenz dann auch häufig in sogenannten Fusionsspielen. So ist in der Praxis von Wargames die sogenannte Fusion von Spielen ein häufig erwähntes Phänomen, nach dem Lernspielen auch analytische Elemente enthalten und analytische Spielen zusätzlich ein Trainingsaspekt innewohnt.

Seine epistemische Struktur positioniert Wargaming insofern auch als eine Untersuchungsmethode (Rubel 2006, 2) – beispielsweise menschlicher Entscheidungsfindungs- und Lernprozesse auf taktischer, strategischer oder ethischer Ebene –, sodass ihr im Vergleich zu anderen Mitteln der Konfliktvorbereitung – wie dem Studium von Geschichte und Theorie, operativer Erfahrung oder der Analyse empirischer Daten – ein einzigartiges Potenzial zugeschrieben wird (vgl. z.B. Lin-Greenberg et al. 2022, 86): Es soll Individuen auf die kommunikativen Prozesse vorbereiten, die sich in Kontexten entfalten, die von Unsicherheit und Komplexität geprägt sind. Diese pädagogische Rahmung wird somit zu einer zentralen theoretischen Position, da sie den Wissenserwerb als zentrales Element des Wargames in den Vordergrund stellt. In Hirsts Lesart (2024, 194–206.) wird der durch ein Wargame hervorgerufene Flow-Zustand aktiv insbesondere dafür genutzt, um Militärangehörigen spezifische Lehren zu vermitteln und ihnen so zu ermöglichen, ihr »mentales und physisches Gedächtnis zu kultivieren« (Übersetzung durch die Verfass.). Vor diesem Hintergrund formuliert Robert C. Rubel (2006, 111), ehemaliger Dekan des Center

for Naval Warfare Studies am Naval War College, die Notwendigkeit von Wargames in der militärischen Ausbildung nahezu kategorisch: »We war-game because we must. There are certain warfare problems that only gaming will illuminate.« Wenn Perla (2016, 173) in diesem Zusammenhang von der Schaffung einer spezifisch *gemeinsamen Erfahrung* spricht – was übrigens auch der Etymologie des Begriffs »Game« entspricht (vgl. Harper 2025a) –, bringt er eine in der Wargaming-Community weit verbreitete Ansicht zum Ausdruck. Danach ist die konkrete Interaktion zwischen zukünftigen Kollaborateur:innen in Konfliktsituationen ein Nebeneffekt der Lernziele situatives Verständnis und Kenntnis von Systemfähigkeiten (vgl. Mouat 2022, 211; vgl. auch Hirst 2024; Rubel 2006). Der pädagogische Wert solcher immersiven und geteilten Erfahrungen gewinnt auch in der akademischen Forschung zunehmend an Bedeutung, wonach synthetische Erfahrungen – bei denen Proband:innen mit immersiven Fiktionen konfrontiert werden – Entscheidungsprozesse auslösen, die mit denen in realen Kontexten vergleichbar sind (Lin-Greenberg et al. 2022, 89).

⇒ 3 Ansprüche an Wargaming im Zeichen der Künstlichen Intelligenz

Der Wunsch, die Komplexität des menschlichen Elements – insbesondere die Muster menschlicher Entscheidungsfindung – zu erforschen, gilt als die treibende Kraft hinter dem anfänglichen Aufstieg von Wargaming im Rahmen der militärischen Ausbildung in den USA in den 1960er Jahren (vgl. Hirst 2024, 57–63.; Emery 2021a). In vergleichbarer Weise liegt dieses Interesse der aktuellen Renaissance von Wargaming seit 2014 zugrunde. Paradoxaerweise geht dieses erneute Interesse jedoch – wie im Folgenden gezeigt wird – in beiden Fällen mit einer parallelen Hinwendung zu Methoden einher, die darauf abzielen das menschliche Element und damit die ihm innewohnende Subjektivität einzudämmen oder zu begrenzen. Dieses Paradoxon illustriert sich besonders deutlich in den jüngsten Bemühungen, große Sprachmodelle in die Wargaming-Praxis zu integrieren.

⇒ 3.1 Wargaming und Automatisierung: Ein Blick zurück

Wie skizziert, schwankt das Interesse an Wargames – als eigenständiger Methode für Bildung und Ausbildung sowie als Instrument im Bereich der internationalen Beziehungen (vgl. Hirst 2024, 58–61) – in militärischen Anwendungen zwischen den Polen Subjektivität und Objektivität. Wie sowohl Hirst (ebd.) als auch John R. Emery (2021a) rekonstruieren, basierte die explizite Hinwendung insbesondere der RAND

Corporation im Kontext des Kalten Krieges der 1960er Jahre zum Wargaming zunächst am Interesse an der Komplexität menschlicher Entscheidungsfindung. Im Laufe der Zeit setzte sich jedoch die an der RAND etablierte Form des quantitativen Wargame – basierend auf computergestützten Simulationen – gegenüber den qualitativen Ansätzen der Social Science Division (SSD) durch und behauptete sich durch ihren Anspruch auf Objektivität (Emery 2021a, b) und den prädiktiven Nutzen von Simulationen (Kaminski et al. 2023, 12).³ Dies geschah trotz der Ergebnisse qualitativer Wargames, die zeigten, wie ein größerer Realismus im Szenariodesign ein hohes Maß an emotionalem Engagement bei den Teilnehmer:innen hervorrief – wodurch diese die Schwere von Entscheidungsprozessen unmittelbar spüren konnten. Doch in dem Maße, wie quantitativ generiertes Wissen – basierend auf Generalisierbarkeit – als die einzig legitime epistemische Form angesehen wurde, wandte sich die Forschung zunehmend von der Förderung der Emotionen und der ethischen Intuition der Spieler:innen (Emery 2021a, 29) ab, die die SSD in ihrer Arbeit über Verhandlungen und diplomatischen Austausch betont hatte. Stattdessen wurden menschliches Urteilsvermögen und Emotionen systematisch im Spielprozess eliminiert, um eine wissenschaftlichere Art der Kriegführung zu schaffen (Emery 2021a, 20), die dem Ideal der mathematischen Abschätzung möglicher Zukünfte (vgl. Emery 2021a, b; Kaminski et al. 2023) folgt.

Nach einer langen Periode, in der das Phänomen Wargaming, mit wenigen Ausnahmen, von der wissenschaftlichen Gemeinschaft weitgehend ignoriert wurde (Hirst 2024; Rubel 2006) und Fragen der Subjektivität – insbesondere hinsichtlich der Validität qualitativer Wargames – auf bestimmte militärische Diskursformate beschränkt blieben (vgl. beispielsweise Debatten um unethisches Wargaming; vgl. Downes-Martin 2013; CGSC 2021), sind zunehmend Untersuchungen auszumachen, die sich mit der Phänomenologie subjektiver Erfahrung auseinandersetzen (vgl. z.B. Zuber 2024) oder systematisch die Gelingensbedingungen qualitativen Wargamings betrachten (vgl. z. B. Jahanbakhsh et al. 2023; Zwarts 2024; Wilden et al. 2023). Hintergrund dieser Entwicklungen ist eine Renaissance des Wargamings – ein erneutes Interesse an dieser Methode im Bereich der internationalen Beziehungen –, die sich mit Hirst (2024, 64–84) auf Veränderungen im US-Verteidigungssektor seit 2014 zurückführen lässt, insbesondere

(3) Nach Sabin (2014, 58–61) ist dieses überwiegende Interesse am quantitativen Interesse schon bis in die 1940er Jahre zurückzuverfolgen.

auf die Defense Innovation Initiative und die Third Offset Strategy. Entscheidend für dieses Wiederaufleben ist die Entwicklung von und Anpassung an neue Technologien, die auf der Arbeit von Institutionen wie der Defense Advanced Research Projects Agency (DARPA) und dem Office of Net Assessment aufbauen. Sie zielen darauf ab, neue strategische Schlüsselbereiche wie etwa Cyber- und elektronische Kriegführung, Mensch-Maschine-Teamarbeit und Wargaming selbst und damit einhergehende neue kognitive und operative Paradigmen zu erschließen: Vor dem Hintergrund einer wahrgenommenen Transformation des Charakters des Krieges – gekennzeichnet durch zunehmende Komplexität und Mehrdeutigkeit – soll Wargaming nicht nur die Grenzen des Möglichen austesten und das Spektrum von operativen Konzepten über Taktik bis hin zur strategischen Planung abdecken. Um die gewünschten operativen und strategischen Effekte zu erzielen, wurde Wargaming zudem als Mittel zur Gestaltung doktrinärer, organisatorischer und strategischer Ansätze für neue Technologien und gleichzeitig als Schlüssel zum Verständnis der menschlichen Dimension angesehen – und damit als Möglichkeit für den Erwerb einer Form des kritischen Denkens, bei dem die Fähigkeit im Vordergrund steht, Komplexität durch gelebte Erfahrung zu bewältigen und die Entscheidungskompetenz in einem neuen Sicherheitsumfeld zu verbessern.

Wenn die Betonung menschlicher Wahrnehmung und Entscheidungsfindung eine Abkehr von den vorherrschenden quantitativen Wargames und deren Bekenntnis zu objektiver Modellierung und Vorhersage signalisierte, so lässt sich in Hirst (2024, 57–89.) Rekonstruktion des Diskurses des Verteidigungsministeriums gleichermaßen ein anhaltender erkenntnistheoretischer Bruch identifizieren, der bis in die 1960er Jahre zurückreicht. Es handelt sich dabei um eine Spaltung innerhalb der Wargaming-Community, die bis heute fortzubestehen scheint: Auf der einen Seite sind diejenigen positioniert, die den subjektiven Charakter von Wargaming und dessen pädagogischen Imperativ, entscheidungsrelevante Intuition zu fördern, betonen. Auf der anderen Seite stehen die Befürworter:innen eines technologischen und datengetriebenen Wargaming-Paradigmas, das Vorhersagefähigkeiten auf der Grundlage von Automatisierung und Berechnung anstrebt. Während die Forderungen nach der Integration Künstlicher Intelligenz in militärische Operationen – und ebenso in Wargames – zunehmen, haben die rasanten Fortschritte im Bereich der Großen Sprachmodelle zugleich eine neue Wahrnehmung hervorgerufen: dass solche Modelle in der Lage sein könnten, menschliche Entscheidungen in einem noch nie dagewesenen Ausmaß und mit noch nie dagewesener Geschwin-

digkeit nachzubilden und mögliche Zukünfte auf eine bisher unerreichbare Weise vorherzusagen.

⇒ 3.2 Wargaming und Automatisierung: Ein Blick in Gegenwart und Zukunft

Die Definition großer Sprachmodelle (LLMs) ist nach wie vor uneinheitlich. Im weitesten Sinne können sie als neuronale Sprachmodelle verstanden werden. Diese zeichnen sich durch eine hohe Anzahl von Parametern aus und werden anhand umfangreicher Textkorpora trainiert, um die statistischen Eigenschaften von Sprache zu erlernen, und darauf basierend probabilistische Vorhersagen über natürliche Sprachsequenzen zu generieren (Weidinger et al. 2022). Mit der weiten Verbreitung von ChatGPT und anderen generativen KI-Modellen – die darauf zielen, auf der Grundlage von Eingabeaufforderungen und Kontextinformationen neuartige Strategien und Entscheidungen zu entwickeln – ist die Aussicht auf die Integration autonomer Agenten in hochriskante Kontexte wie militärische und diplomatische Entscheidungsfindung zunehmend plausibel geworden (vgl. z.B. Hoffman und Kim 2023). Nimmt man als Ausgangspunkt die idealisierte Struktur eines Wargames – in dem die Teilnehmer:innen Ziele definieren, die auf szenariospezifischen Siegbedingungen der Auftraggeber:innen und Designer:innen des Spiels abgestimmt sind, Ressourcen strategisch verteilen, von Spielleiter:innen (Facilitators) und möglicherweise Schiedsrichter:innen, sogenannten Adjudicators, durch das Spiel geführt werden und sich an einer Nachbewertung beteiligen – dann, so die Vision, könnten LLMs auf nahezu jeder Ebene des Wargames eingesetzt werden, die einst als Domäne menschlichen Handelns galt (für einen technisch orientierten Überblick vgl. Gallotta et al. 2024). Sie sollen die Funktion übernehmen, Szenarien zu entwerfen und/oder zu simulieren (Davis und Bracken 2025; Chen und Chu 2024), als Facilitators, Instruktoren oder Adjudicators fungieren (vgl. Hogan und Brennen 2024; Fight Club International 2025) oder das Gameplay analysieren (Seater 2021). Nicht zuletzt wird der »Homo Silicus« (Horton 2023) visioniert – eine projizierte Entität, die den Menschen als Spieler:in oder Gegner:in vollständig ersetzen soll (vgl. Hogan und Brennen 2024; Liu et al. 2024). Daraus wird geschlossen, dass »LLMs may supplant human participants for data collection.« (Grossmann et al. 2023, 1108)

Der aktuelle Hype um die Integration von LLMs in Wargames wird wesentlich durch die dramatische Zunahme der Rechenkapazität aufrechterhalten, die die Entwicklung zahlreicher anderer Technologien im Bereich der Künstlichen Intelligenz vorangetrieben hat. Eine dieser

Fähigkeiten ist die Gewinnung von Erkenntnissen aus riesigen Datensätzen in kürzester Zeit (Davis und Bracken 2025). Angesichts der Analysegeschwindigkeit und der dokumentierten Fähigkeiten von LLMs in strategischen Spielen (vgl. z.B. Silver et al. 2016; FAIR 2022; Goodman et al. 2020) sehen ihre Befürworter:innen eine Zukunft, in der Wargames dann auch in schnellen, groß angelegten Iterationen durchgeführt werden können (Shrivastava 2024) – nicht nur um Kosten zu senken oder Zeit zu sparen (Chen und Chu 2024), sondern auch um die logistischen Herausforderungen zu mildern, die mit der Entwicklung und Durchführung von Wargames verbunden sind. Dazu gehören der hohe Zeitaufwand der Szenariomodellierung und ein Mangel an Fachexpert:innen, die als Instructors und Adjudicators fungieren können (Weller et al. 2024). Durch maschinelles Lernen – insbesondere Mustererkennung – wird LLMs zudem das Potenzial zugeschrieben, den Ausgang von Wargames selbst vorherzusagen oder gegnerisches Verhalten zu antizipieren (vgl. z.B. Zhang et al. 2024).

Während diese Versprechen in erster Linie auf die vollständige Substitution menschlichen Handelns abzielen, fokussiert ein zweites zentrales Anwendungsszenario die Nutzung von Sprachmodellen, um menschliche Entscheidungen in Form von Chatbots zu unterstützen. Sie sollen auf der Basis von Texteingaben begründete Ergebnisse in natürlicher Sprache liefern (vgl. Chen und Chu 2024; Hogan und Brennen 2024; Rivera et al. 2024; Mitsea et al. 2025). Der Reiz dieser Spielerassistenz – Agenten, die mit der Spielumgebung auf einem Niveau interagieren könnten, das dem menschlicher Spieler:innen nahekäme – liegt vor allem in einer erwarteten kognitiven Entlastung. Dies wiederum nährt die Hoffnung auf eine Demokratisierung von Wargaming, wodurch auch unerfahrene Teilnehmer:innen leichter Zugang zu den oft durch komplexe Szenarien und komplizierten, nicht einfach zu erlernenden Regelsystemen charakterisierten Spielen erhalten würden (vgl. Thompson 2024).

Insgesamt bleiben die prognostizierten Vorteile der Assistenz jedoch weitgehend dieselben wie sie dem Ersatz menschlicher Akteure zugrunde liegen: erwartet wird menschliche Fehler, Voreingenommenheit und Emotionalität zu mildern und die sogenannte menschliche kognitive Fehlbarkeit im Entscheidungsprozess einzuschränken (vgl. Chen und Chu 2024; Helberger et al. 2020). Argumente für ein solches »superhuman play« (sic!) (Shrivastava 2024, 2) stützen sich oft auf die vermutete Fähigkeit von LLMs dem Menschen gleich zu handeln: aus dem sich wiederum die Erwartung speist, »simulating human-like responses and behaviors offers opportunities to test theories and hypotheses

about human behavior at great scale and speed« (Grossmann et al. 2023, 1108; vgl. auch Chen und Chu 2024, 2895). Das besondere Potenzial von LLMs wird insofern darin gesehen, menschenähnliche Reaktionen auf sprachliche Eingaben zu generieren und dadurch Entscheidungsprozesse zu verbessern. Indem sie intuitive, vielseitige und kontextreiche Unterstützung durch die Kombination natürlicher Sprachverarbeitung mit fortschrittlichen Analysefunktionen böten, sollen die menschlichen Fähigkeiten in einem breiten Spektrum von Entscheidungskontexten erweitert und verstärkt werden (Chen und Chu 2024, 2895).

⇒ 4 Risiken der Kombination von Künstlicher Intelligenz und Wargaming

Die bisher dargelegten Argumente für die Integration von LLMs in Wargames sind nicht unumstritten. Es lassen sich eine Reihe kritischer Einwände formulieren – beispielsweise, ob das oft angeführte Argument der Datenknappheit auch im Sinne eines Fehlens von »subject matter experts«, das den Einsatz solcher Technologien vorgeblich rechtfertigt, nicht in Wirklichkeit eine grundlegendere Frage verdeckt: nämlich, warum Daten – insbesondere einer bestimmten Art – überhaupt fehlen oder unzureichend sind (Agnew et al. 2024). Hier ist z. B. an die spezifische Professionalisierung des Wargaming durch eine mehr oder weniger geschlossene Wargame-Community (vgl. Hirst 2024; Sabin 2014) zu denken, die ihr Anliegen über mehrere Jahrzehnte gegenüber dem Objektivierungsanspruch von computergenerierten Simulationen verteidigen musste. Mit anderen Worten: Könnte nicht gerade der relative Mangel an empirischer Forschung zu subjektiven Faktoren in militärischen Entscheidungsprozessen, die im Wargaming gefördert werden sollen, selbst eine Folge eines epistemischen Bekenntnisses zur Objektivierung sein? Und untergraben neuere technologische Entwicklungen und ihre Implementierung in Wargaming-Kontexten letztlich die neuerlich privilegierte Betonung menschlichen (Experten-)Wissens? Diese Frage erhält eine qualitativ neue Dimension, wenn LLMs nicht mehr nur der schnellen Informationsauswahl dienen, sondern, weil sie zunehmend adaptives Reaktionsverhalten zeigen, zu aktiven Teilnehmerinnen an Sinngebungsprozessen werden sollen.

So wurden etwa mit Blick auf die Abstraktion des Spiels von einer tatsächlichen Konfliktsituation immer schon Bedenken hinsichtlich möglicher Verzerrungen und der Entstehung spielspezifischer Artefakte geäußert (Downes-Martin 2013; CGSC 2021), die sich als Risiko eska-

lierender Tendenzen einzelner Spielzüge darstellen. So argumentiert Rubel (2006, 116), wenn auch ohne empirische Untermauerung, dass Spieler:innen in simulierten Kontexten aggressiver agieren könnten als in realen Situationen. Ist dieses Problem in der Literatur vor allem als Designimperativ formuliert, nämlich Spiele so realitätsnah zu strukturieren, dass die Wahrscheinlichkeit und Intensität solcher Artefakte minimiert werden (ebd.), stellt es nicht nur einen wesentlichen Einwand gegenüber den seit den 1960er Jahren dominierenden quantitativen Wargames und ihrem Anspruch auf Objektivität dar. Er bildet auch heute der Kern einer Kritik an der (Teil-)Automatisierung von Wargames. So zeigen empirische Studien, dass LLMs, eingesetzt als eigenständige Spielakteure, überproportional militarisiertere oder aggressive Aktionen bevorzugen, selbst wenn das Szenario oder die Spielparameter differenziertere oder defensivere Strategien nahelegen würden. Rivera et al. (2024) beobachteten beispielsweise, dass alle fünf der von ihnen untersuchten Modelle eine Neigung zu Rüstungswettläufen zeigen, die mitunter zu groß angelegten Konflikten und in seltenen Fällen sogar zu Atomkriegen eskalierten. Eine qualitative Analyse der von diesen Modellen gelieferten Argumentation für ihre gewählten Handlungsweisen ergab eine Abhängigkeit von Abschreckungs- und Erstschlaglogik. Während einige Autor:innen solche Ergebnisse auf spezifische Merkmale des Spieldesigns zurückführen – wie Spielmechanik, Szenariogestaltung oder Belohnungsstrukturen, aufgrund derer die Modelle in Richtung militarisierter Optionen tendieren könnten (vgl. z.B. Weller et al. 2024, 9) –, vermuten andere Korrelationen zwischen der politischen Haltung der durch das Modell repräsentierten Nation oder Persönlichkeit und den daraus resultierenden Entscheidungen (Hogan und Brennen 2024, 9). Lamparth et al. (2024) kommen dagegen zu dem Schluss, dass solche Eskalationstendenzen in erster Linie auf die eingeschränkten Echtzeit-Lernfähigkeiten von LLMs zurückzuführen sind, was auch Folgen für den Einsatz von Großen Sprachmodellen in Form von Assistenzsystemen hat. So sind die Bedingungen, unter denen eine solche »artificial escalation« (Aguirre et al. 2023) entsteht, teilweise auf die Priorisierung relativer quantitativer Kennzahlen durch KI-Systeme zurückzuführen – anstelle der qualitativen Bewertung absoluter Ergebnisse (vgl. Davis und Bracken 2025).

Eine Beteiligung von LLMs am Wargaming deutet dann auch nicht nur auf eine Veränderung in der Art und Weise hin, wie Informationen verarbeitet werden, sondern auch auf eine Veränderung, wie Realitäten erzeugt und stabilisiert werden – wobei LLMs eine mitgestaltende Rolle übernehmen. Die mit dem Einsatz von LLMs in Wargames verbun-

denen Risiken, wie sie weiter in den folgenden Abschnitten erörtert werden, beinhalten daher nicht nur technische und mithin praktische oder operative, sondern auch erkenntnistheoretische und ethische Überlegungen. Diese werden besonders deutlich, wenn man die Natur der technischen Handlungsfähigkeit von LLMs betrachtet – in erster Linie als Erweiterung der menschlichen Wahrnehmung und anschließend als autonome oder halbautonome unterstützende Kraft innerhalb von Entscheidungsarchitekturen.

⇒ 4.1 Technisch-funktionale Einwände

Die Argumentation, LLMs in das Wargaming zu integrieren, verweist in erster Linie auf die jüngsten Fortschritte, die LLMs als überlegen gegenüber traditionellen KI-Methoden in der Entscheidungsfindung ausweisen: Danach zeichnen sie sich durch die Fähigkeit aus, komplexe, differenzierte Daten zu interpretieren, über atypische Szenarien hinweg zu verallgemeinern, womit sie ihre Befähigung menschliche kognitive Prozesse widerzuspiegeln, unter Beweis stellten (Chen und Chu 2024, 2896). Die Forschung illustriert demgegenüber aber auch, dass LLMs zwar strategisch spielen können, aber weder die Dynamik eines Spiels noch dessen Ergebnisse in gleichem Maße wie es menschliche Spieler:innen tun, qualitativ zu bewerten in der Lage sind. Am Spiel assistiver LLMs für Spieler:innen veranschaulichen Yu et al. (2023) und McKinnon et al. (2022) etwa ihre Anfälligkeit für strategisch formulierte – auch kontroverse – Eingabeaufforderungen und stützen damit die Schlussfolgerung von Lamparth et al. (2024), dass LLMs neue Eingaben mitunter nicht richtig interpretieren. Diese Befunde stehen im Einklang mit den Ergebnissen einer Evaluation der LLM-Leistung in einem Matrix-Wargame des International Fight Club (2025). In dieser Art von Entscheidungsspiel, das typischerweise in einem Seminarumfeld gespielt wird, schlagen die Spieler:innen Aktionen vor, deren Erfolg durch argumentativen Austausch auf der Grundlage sachlicher, umsetzbarer Behauptungen bestimmt wird, die in einer moderierten oder konsensorientierten Diskussion geklärt werden (Curry et al. 2020). Das im Matrix-Spiel als Assistent eingesetzte LLM zeigte jedoch Inkonsistenzen bei der Verfolgung des Fraktionseinflusses und der öffentlichen Stimmung über mehrere Runden hinweg und führte zu Schlussfolgerungen, die nicht vollständig mit den etablierten Spielereignissen übereinstimmten. Die Evaluator:innen kamen zu dem Schluss, dass menschliche Spieler:innen ihre Positionen im Laufe des Spiels dynamisch anpassen, das verwendete LLM diese sich entwickelnden Positionen jedoch nicht konsistent widerspiegeln konnte. Vergleichbar

dokumentierte Shrivastava (2024, 11) Diskrepanzen in den Antworten von LLMs und stellte fest, dass semantisch äquivalente Eingabeaufforderungen in zwei Spielszenarien zu deutlich abweichenden Empfehlungen führten – was einen Mangel an innerer Kohärenz offenbarte. Ein solcher Mangel an Kohärenz ist dann damit verknüpft, dass LLMs lediglich menschliches Sprachverhalten simulieren, wie unter anderem Lamparth et al. (2024) argumentieren. Sie (ebd.) betonen zudem, dass Sprachmodelle, da sie im Kern algorithmische Systeme sind, die aus Code bestehen, so konzipiert sind, dass sie auf Eingabeaufforderungen in vorprogrammierter Weise reagieren – mit anderen Worten: LLMs führen keine Denkprozesse durch (vgl. z.B. Felder und Kückelhaus 2025, 446). Weitere Befunde stützen diese Charakterisierung. Lamparth et al. (2024) beobachten zudem, dass simulierte Spieler:innen typischerweise nur kurze Stellungnahmen abgeben, selten auf Meinungsverschiedenheiten eingehen und dazu neigen, eine bevorzugte Vorgehensweise zu benennen, während sie hierfür nur begrenzte Argumente liefern und sich minimal auf vorherige Diskurse beziehen. Dies wiederum lässt sich in Kontrast zur Verhandlungsdynamik des menschlichen Spiels setzen, in dem gegnerische Teams politisch-militärische Signale austauschen, um abzuschätzen, was die andere Seite akzeptieren oder ablehnen könnte. Laut Lin-Greenberg et al. (2022, 90–91.) sind solche Interaktionen zudem nicht rein rationale Berechnungen; sie beinhalten vielmehr Prozesse des Abwägens konkurrierender Prioritäten unter Bedingungen der Unsicherheit – eine zudem affektive Dimension, die für die Relevanz von Wargaming zentral ist (Hirst 2024, 194–198), für LLMs in dieser Hinsicht jedoch unzugänglich bleibt.

Große Sprachmodelle treffen ihre Entscheidungen – sieht man von ihrer Orientierung am Gewinn des Szenarios ab – ohne Ziel bzw. ist es lediglich »ihr Ziel, je nach Input das vorherzusagen, was laut ihren Trainingsdaten am wahrscheinlichsten ist« (Esposito 2024, 62). Wie die oben genannten Studien veranschaulichen, kann dieser Prozess bei identischen oder semantisch äquivalenten Eingaben zu unterschiedlichen Ergebnissen führen und zudem, was das rationale Schlussfolgern betrifft, den Bezug zu den Eingaben in das Modell verlieren – was weit hin unter dem Begriff des Halluzinierens diskutiert wird. Dabei handelt es sich um einen Terminus, der – die vorherrschende Anthropomorphisierung von KI-Technologien widerspiegelnd – die Tendenz des Modells bezeichnet, von der Faktizität abweichende Inhalte zu generieren, die zwar kohärent und glaubwürdig erscheinen, in Wirklichkeit aber erfunden oder unbegründet sind (Lin et al. 2024). Diese Tendenz wurde ebenfalls in der Evaluation des Fight Clubs (2025) sichtbar, in der das

Sprachmodell Argumente produzierte, die von den Absichten der Spieler:innen abwichen. In diesen Befunden illustriert sich das epistemische Problem der falschen Darstellung und insbesondere der Desinformation (vgl. Agnew et al. 2024). Hogan und Brennen (2024, 6), die dieses Phänomen in ihrer Studie ebenfalls beobachteten, interpretieren die LLM-Halluzination gleichwohl als epistemischen Vorteil im Kontext von Wargames. Für sie bildet das Halluzinieren ein Phänomen einer »well-grounded creativity required to generate plans and adjudicate their outcomes is hallucination by another name«. Mit dem Bezug des Halluzinierens auf Kreativität verweisen sie auf ein schöpferisches Moment und mithin eine Fähigkeit, die zwar als wesentlich für die Play-Dimension von Wargaming gilt (vgl. Hirst 2024, 143). Diese, gleichermaßen kreative, Interpretation ist jedoch mit Vorsicht zu genießen. Wie vielfach betont und bereits aufgeführt, kennen LLMs keine Fakten wie Menschen. Sie generieren Texte, indem sie das nächste Token auf der Grundlage von Wahrscheinlichkeiten vorhersagen und nicht durch Schlussfolgerungen oder die Überprüfung der Wahrheit. Während David Chapman (2023, 195) LLMs vor diesem Hintergrund passend als »text genre imitation engines« charakterisiert und darin ebenfalls ein gewisses kreatives Potenzial erkennt, wäre es allerdings irreführend, daraus auf einen Anschein kognitiver Prozesse zu schließen. Halluzinationen sind in erster Linie eine Funktion fehlerhafter Trainingsdaten, der Architektur des Modells und/ oder der prozeduralen Dynamik, durch die Inhalte generiert werden (vgl. Lin et al. 2024 für einen Überblick): In manchen Fällen können die Daten zu spezifisch oder zu allgemein sein – und vor allem sind komplexe Szenarien, wie sie für viele Wargames von zentraler Bedeutung sind, für LLMs nach wie vor besonders schwierig zu erfassen (Chen und Chu 2024, 2895).

Die bis hierhin skizzierten technischen Einwände werfen in erster Linie das Problem der Verantwortbarkeit auf, und dies zunächst hinsichtlich der Richtigkeit der Ergebnisse eines Modells. Die inhärente Komplexität von LLMs – die selbst für Expert:innen als mitunter schwer zu erfassen gilt – macht Fragen nach ihrer Transparenz und Interpretierbarkeit daher besonders dringlich (vgl. Goodman et al. 2020; Black und Darken 2024). Licato (2022) argumentiert deshalb, dass der Weg, auf dem Argumente zu einer endgültigen Schlussfolgerung integriert werden, im Sinne einer »argument-justified AI« verständlich gemacht werden muss, ohne die die Validierung und der ethische Einsatz KI-generierter Strategien zutiefst problematisch bleiben.

⇒ 4.2 Epistemisch-moralische Einwände

Im militärischen Kontext gilt schnelle und präzise Entscheidungsfindung als entscheidende Kompetenz. Erst sie ermögliche eine effiziente und logisch schlüssige Durchführung von Operationen angesichts der inhärenten Komplexität und Unvorhersehbarkeit zukünftiger Schlachtfelder und Konfliktverläufe. Das Training dieser Fähigkeit ist ein zentrales Ziel von Wargaming. Der Einsatz von Automatisierungstechnologien wie Large Language Models als Entscheidungsunterstützungssysteme zur Bewältigung solch vielschichtiger Herausforderungen birgt jedoch Risiken, die das Prinzip fundierter Entscheidungsfindung selbst beeinträchtigen. Shrivastava (2024), dessen Studie Inkonsistenzen im Reaktionsverhalten verschiedener LLMs aufzeigte, weist etwa auf das trügerische Vertrauen hin, das ihre Nutzer:innen in die Ergebnisse der Modelle setzen könnten (ebd., 11) – ein Vertrauen, das menschliche Entscheidungsträger:innen in die Irre führen kann. Dieser Besorgnis liegt das Phänomen des sogenannten digitalen Animismus (Moos 2021) zugrunde, der die Aufmerksamkeit auf die von LLMs eingesetzten linguistischen Strategien lenkt – nämlich ihre Simulation zwischenmenschlicher Zugänglichkeit. Wie Felder und Kückelhaus (2025, 431) argumentieren, führen diese Strategien »zu einer Verschmelzung von Realität und medial vermittelter Realität, die für Nutzer:innen potenziell täuschend sein kann«. Die Anthropomorphisierung von LLMs – d. h. die Tendenz, ihnen Bewusstsein oder Intentionalität zuzuschreiben – gilt dabei nicht nur als ein kognitiver Fehler, sondern vielmehr als eine kulturell verankerte anthropozentrische Disposition, die, wie Untersuchungen nahelegen (vgl. Thompson 2024; Weidinger et al. 2022), das Vertrauen der Nutzer:innen in das LLM und seine Ergebnisse deutlich erhöht. Dementsprechend werden LLMs, die als Co-Piloten menschliche Spieler:innen unterstützen, wahrscheinlich nicht nur die Modalitäten der Mensch-Maschine-Interaktion durch Echtzeit-Anleitung neugestalten, sondern sind auch geeignet die Autonomie der Spielerentscheidungen zu beeinträchtigen. Wenn sich Spieler:innen zunehmend auf KI-generierte Empfehlungen verlassen, laufen sie Gefahr, sich von ihren eigenen, unabhängigen kognitiven und affektiven Fähigkeiten zu lösen. Dies weist auf ein weiteres Problem hin: Die Glaubwürdigkeit, die LLMs durch flüssige und zugängliche Ergebnisse vermitteln, birgt nicht nur das Risiko übermäßiger Nutzervertrauens, sondern fördert auch den Eindruck von Objektivität. Indem ihr Einsatz darauf zielt, Subjektivität explizit auszuschließen und große Datensätze schneller verarbeiten, als es die menschliche Wahrnehmung zulässt, verspricht er – dann auch in ihrem Einsatz als Spieler:in

selbst – implizit objektives Wissen über »richtige« Strategien und wünschenswerte Ergebnisse zu produzieren. Doch Davis und Bracken (2025, 35) stellen zu Recht die Frage:

what good is a database of a million scenarios if the underlying models assume perfect rationality, perceptions, alliance relationships, and focus on, say, the post-exchange ratio of nuclear weapons as a measure of outcome?

In diesem Sinne steht das Streben nach Objektivität – und die Automatisierungsgetriebene Annahme, dass wiederholte Iterationen verallgemeinerbare Ergebnisse liefern können – auch im Widerspruch zu den pädagogischen Zielstellungen traditioneller Wargames und stellt mithin das Konzept des Wargaming faktisch auf den Kopf. Tatsächlich besteht das Risiko nicht nur darin, dass Spieler:innen in Zukunft die Fähigkeit verlieren, fundierte Entscheidungen zu treffen – Fähigkeiten, die sie andernfalls durch wiederholte Auseinandersetzung mit Entscheidungsprozessen entwickelt hätten –, sondern auch darin, dass nicht nur der kognitive, sondern auch affektive Prozess der Entscheidungsfindung selbst teilweise externalisiert und ausgelagert wird resp. letzterer nicht zur Entfaltung gelangt. Dies ist Gegenstand der nachfolgenden Ausführungen, denn aus einer technikethischen Perspektive ist mit der Integration von Sprachmodellen in das Wargaming der angeführte Begriff der Verantwortung auch auf das Risiko einer kognitiven Distanz durch algorithmisch verdichtete Entscheidungen resp. Empfehlungen zu beziehen. So ist die Frage nach dem verantwortungsvollen Umgang mit (neuen) Technologien, wie etwa Gerhard Schreiber (2025) betont, immer auch mit jener danach verknüpft, wie wir auf etwas reagieren, das uns widerfährt. Sie ist demnach eine Auseinandersetzung mit etwas, das uns betrifft, nämlich wie Technologien unser (menschliches) Handeln im Sinne von Fähigkeiten, Fertigkeiten und Kompetenzen beeinflussen.

⇒ 4.2.1 Technisiertes Wargaming und das Risiko des De-Skilling

Vor diesem Hintergrund können die Ergebnisse der referierten technischen Studien als Hinweis auf die mit der Automatisierung im militärischen Kontext verbundenen Risiken gelesen werden: Spieler:innen werden mit den Konsequenzen von Entscheidungen konfrontiert, die primär von der »Maschine« – und erst sekundär vom Menschen – getroffen werden. Wenn man, wie dieser Beitrag, argumentiert, dass

LLMs im Wargaming genau den Erfahrungsraum einschränken, den diese Spiele eigentlich eröffnen sollen, dann rührt diese Einschränkung aus der paradoxen Spannung zwischen synthetischer und immersiver Erfahrung.

Das Spielszenario eines Wargames und seine Regeln sind, wie ein literarischer Text, Fiktionen, und insofern nicht per se von sich aus bedeutungsvoll. Wie der Literaturwissenschaftler Wolfgang Iser (1976) entfaltet, entsteht die Bedeutung eines literarischen Werks durch die Interaktion zwischen Leser:in und Text. Da Texte Lücken oder Unbestimmtheiten enthalten, müssen seine Leser:innen diese füllen, um Bedeutung zu konstruieren. Überträgt man diese Lesart auf das Wargaming, dann ist zwar einschränkend zu bemerken, dass Wargames, anders als Gedichte oder Belletristik, keine lineare Struktur aufweisen. Hirst (2024, 195–196.) spricht in diesem Zusammenhang (mit Bezug auf James Dunnigan) etwa von Wargames als »non-linear communication devices«. Gleichwohl können Wargames als ein Medium verstanden werden, das Elemente der vorweggenommenen militärischen Realität auf eine Weise organisiert, darstellt und operationalisiert, die Bedeutung erzeugt resp. gelesen und gedeutet wird.

Wargames konstituieren sich, mit anderen Worten, durch eine spezifische Erzählung, denn in ihnen werden bestimmte Werte und Ideen gleichermaßen betont, wie andere Themen, Aspekte der Geschichte oder Prinzipien versteckt werden (Hirst 2024, 196). Die Bedeutung des Rezeptionsaktes im Wargame durch die Spieler:innen ist daher darin zu sehen, was Hirst unter der Lupe des spezifischen Subjektivierungsmodus des Wargaming analysiert, der durch die Autorenschaft des Spiels geleistet wird: Designer:innen, im Auftrag spezifischer Sponsor:innen und nach deren Vorgaben, entscheiden über Spielziele und übermitteln über Struktur und Abläufe des Spiels bestimmte Informationen – »messages, concerns, and even conclusion« (Perla 1990, 185). Ist die Vermittlung der jeweiligen militärischen Doktrin hierin angelegt, verweist dies wiederum auf die besondere Bedeutung, die Wargames seit 2014 zugeschrieben wird. Im Kontext der 3OS, die, in ihrem Bezug auf die Verwobenheit aktueller verteidigungspolitischer Herausforderungen und der Relevanz neuer Technologien, explizit der Status von Wargaming als »Quelle doktrinärer Innovation« (Hirst 2024, 75) herausstellte, wurde Wargaming hier als Möglichkeit adressiert, die damit einhergehenden komplexen Herausforderungen antizipativ zu meistern, weil sie dazu geeignet seien »[to] provide players with the opportunity to make critical mistakes and learn from

them— and to perhaps reveal breakthrough strategies and tactics when doing so.« (Work and Selva 2015 zit. in Hirst 2024, 75)

An dieser Stelle wird mit dem Eintritt von Technologien wie Großen Sprachmodellen in das Wargaming die »authored agency« (Sicart 2016, 312)⁴, die Spielen zugrunde liegt, virulent – ein Begriff, der auf die klare Absicht verweist, mit der Einschränkungen der Handlungsfähigkeit der Spieler:innen durch Auftraggeber:innen, Designer:innen oder auch Adjudators, wenn diese etwa über den Ausgang von Spieldzügen entscheiden, vorgenommen werden. In Hirsts (2024, 195–211.) dekonstruktiver Lesart des Wargamings führt dies schon bei vom Menschen entworfenen Spielszenarien zu einer Einschränkung des eigentlich betonten »kritischen Denkens« seiner Spieler:innen, welches das Wargaming wiederum ermöglichen soll: die im Spiel angelegte Doktrin ist etwa statt zu befragen zu internalisieren. Übernehmen Sprachmodelle nun aber den Entwurf von Spielen (vgl. z.B. app.pytho.ai), dann bedeutet dies, wie Elena Esposito (2024, 89) anmerkt, nicht nur mit einer Form der Kommunikation konfrontiert zu sein, die keinen Autor, keine Autorin mehr hat. Danach überwindet generative KI eine weitere Schwelle der Unwahrscheinlichkeit, denn »[s]ie ermöglicht die Teilnahme an Kommunikation, ohne mit einem Menschen zu kommunizieren«. Insofern, und gerade weil Krieg sich zunehmend mehr und schneller in technischen Dimensionen vollzieht und das sich darauf beziehende Handeln Teil der militärisch antizipierten Komplexität ist, ist dann auch danach zu fragen, ob ein solch technisiertes Wargaming den ursprünglichen Anspruch, der mit seiner Renaissance verbunden war, letztlich konterkariert. Wie es der Spielethiker Miguel Sicart (2020, 15) formuliert, ist Spielen eine sorgfältig ausbalancierte Aktivität, die eine Welt vorschlägt, ihr durch Regeln und Requisiten Konsistenz verleiht und den Akteuren die Verantwortung überträgt, diese Welt am Leben zu erhalten. Wenn LLMs Wargame-Szenarien kreieren, die also immer auch ein festgelegtes System von Prinzipien, Ansichten und (auch ethischen) Richtlinien umfassen, die zu vermitteln den Kern des pädagogischen Imperativs von Wargames bildet, wessen Ziele und Regeln vermitteln diese Szenarien dann? Und wie, so lässt sich auch vor dem Hintergrund zunehmender Kooperationen mit kommerziellen Innovator:innen, die diese Funktionalitäten für den Spieleentwurf anbieten, weiterfragen, ist sicherzustellen, dass sie auch Regeln für angemessenes Verhalten abbilden?

(4) Noch 2014 hat Sicart die Rolle von Spieledesigner:innen als Autor:innen abgelehnt (90).

Wie schon etwa Shannon Vallor (2015) und in jüngerer Zeit z.B. Sigurd N. Hovd (2023) argumentieren, birgt die Einführung von Technologien im Feld der KI insofern die Gefahr, den Menschen so weit von der Realität des Krieges zu entfernen, dass Fähigkeiten im Sinne eines moralischen Rüstzeugs, die dieser erfordert, nicht mehr entwickelt werden können. Mit diesen theoretischen Einwänden lässt sich vor dem Hintergrund des sich mit Wargaming verbindenden Anspruch mit Sicart (2020, 14) argumentieren, dass sich die menschlichen Akteure in der Infosphäre [...] als Homo poieticus verhalten [sollten], als Verwalter der Umgebung, in der sie gedeihen. In diesem Sinne, Floridi (1999) zitierend formuliert er:

Homo poieticus is a creative, moral agent who inhabits the infosphere, an environment ›constituted by the totality of information entities, including all agents – processes, their properties and mutual relations‹ (Floridi 1999).

Ist insofern bis hierin von der Gefahr des De-Skillings die Rede, dann ist, wie bereits angedeutet, damit gewiss auch eine Einrede in das automatisierte Wargaming selbst formuliert, das immer auch mit dem Vorwurf der kognitiven Distanzierung und Desensibilisierung konfrontiert ist. Vor dem Hintergrund technologischer Umbrüche im Militär illustriert die Vision technische Agenten in das Wargaming zu integrieren aber auch, dass das Risiko der Entfähigung nicht nur die einzelnen Soldat:innen betrifft, sondern die gesamte Institution Militär und insofern das institutionell vermittelte Umfeld, in das das Militärpersonal eingebettet ist, adressiert.

⇒ 4.2.2 Technische versus moralische Agency

Unabhängig davon, ob die Integration von LLMs in Wargames zu einer übermäßigen Abhängigkeit von KI-basierten Werkzeugen in Entscheidungsszenarien führen könnte – in denen seine Nutzer:innen auf Künstliche Intelligenz zurückgreifen, um komplexe Aufgaben zu bewältigen –, stellt die Einbeziehung solcher Technologien das Konzept des Wargames faktisch auf den Kopf. So könnten Wargames, die durch LLMs vermittelt werden, weitgehend auf der Ebene synthetischer Daten angesiedelt bleiben, anstatt ein immersives Erlebnis zu bieten, bei dem emotionale Beteiligung eine entscheidende Rolle spielt – insbesondere im Hinblick auf die Abwägung von Entscheidungen gegen unsichere Ergebnisse. Dies betrifft in jedem Fall ihre Modellierung als eigenständiger Spielereinheit, erweist sich jedoch auch als Gefahr in

ihrer Form als Assistenzsystem, da sie in das Momentum des play eingreifen, denn

[i]f play has a central quality, it is that this behaviour (as action, interaction, and activity), first of all, celebrates people's ability to craft their own responses to circumstances free from *interference*. (Hendricks 2015, 221 zit. in Hirst 2024, 143, Hervh. durch die Verfass.)

Im Lichte des erklärten Zwecks von LLMs im Wargaming menschliche kognitive Fähigkeiten für komplexere und kreativere Aufgaben freizusetzen, könnte man hier von einer solchen »Störung« sprechen und argumentieren, dass Wargaming zu einer bloßen Simulation verkommt, solange das Entscheidungsverhalten des LLMs und die Reaktion auf dieses Entscheidungsverhalten selbst nicht zum Gegenstand des Wargames wird. Ohne eine Reflexion darüber, dass Technologien wie Große Sprachmodelle nicht nur die physische, sondern auch konzeptuelle Umwelt beeinflussen, nähme eine solche Entwicklung den Weg ihrer umfassenderen Vision, nämlich eine vermeintlich fehlerbehaftete Emotionalität der Spieler:innen durch rationales Kalkül zu ersetzen – mit den erwähnten potenziellen Konsequenzen. Unbeachtet bleibe dabei aber auch, dass sogenannte KI-Technologien ein anderes Spiel spielen als Menschen und sich die vielbeachtete Frage, wie Maschinen dem Menschen ähnlich werden können, in eine verwandelt, in der der Mensch in seinem Handeln der Maschine ähneln soll (vgl. Collins 2018, 83–85). Wie oben dargelegt, weisen aktuelle Große Sprachmodelle noch immer erhebliche Defizite auf, die Zweifel an ihrer Fähigkeit aufkommen lassen, durch sprachliche Vermittlung sinnvolle und kontextrelevante Interpretationen zu generieren. Zudem wägen Menschen, im Gegensatz zu Algorithmen, Wahrscheinlichkeiten nicht ausschließlich objektiv ab (Davis und Bracken 2025, 35). Mit anderen Worten – und wie schon länger in Bezug auf computergestütztes Wargaming argumentiert wird (vgl. Emery 2021a) – sind LLMs zwar in der Lage, strategisch zu spielen, aber sie können Ergebnisse nicht in der gleichen Weise qualitativ bewerten wie menschliche Spieler:innen. Dies ist jedoch wiederum entscheidend für die pädagogischen Ansprüche an Wargames, die sich nicht auf die Erforschung taktischer Optionen beschränken, sondern darauf abzielen, emotionales Engagement zu wecken und die »mentale Resilienz« militärischer Akteure zu fördern – wie beispielsweise im Wargaming-Handbuch der Bundeswehr (Doktrinzentrum 2024, 16) dargelegt.

Zudem, und wie Kaminski et al. (2023) konstatieren, basiert die Grundlage der Simulation – ihre Betonung auf Komplexitätsreduktion, Optionsvergleich und politische Intervention – oft auf stark vereinfachten systemischen Funktionen, Ausgangsannahmen und linearen Kausalbeziehungen. Was Simulationen routinemäßig vernachlässigen, sind die singulären Effekte und Kontingenzen menschlichen Handelns. Sie verschleiern zugrunde liegende Unsicherheiten und suggerieren einen Grad an Genauigkeit, den die Realität nur selten, wenn überhaupt bietet. In ihrer Untersuchung, ob die internationalen Beziehungen von Wargaming-Methoden profitieren können, kommen etwa Lin-Greenberg et al. (2022, 87) dann auch zu dem Schluss, dass »interaction between participants better represents real-world decision-making, [...] wargames present players with the consequences of their own decisions«. Die Fähigkeit zu erklären, warum und wie ein bestimmtes Spielergebnis zustande gekommen ist, ist somit nicht nur durch die Intransparenz von LLMs erschwert, sondern diese verändert auch den sozio-erfahrungsbezogenen Bereich – der sowohl explizite (Regeln) als auch implizite (kognitive und emotionale Beteiligung) Elemente umfasst –, den Wargames wiederum greifbar machen wollen. Spielen erfordert ein ständiges Bewusstsein für das Spiel des Gegners und der Gegnerin (und in manchen Fällen auch der Teamkolleg:innen) und eine Reaktion darauf. Angesichts der Intransparenz von LLMs erscheint es nicht möglich, die Strategien oder Reaktionen des Gegners und der Gegnerin intuitiv zu erfassen, weil man eben seine und ihre Spielweise nicht (er)kennt (Coeckelbergh 2022, 153). Mark Coeckelbergh (2022, 152) zieht in diesem Zusammenhang eine überzeugende Analogie, wenn er argumentiert, dass »technologies are embedded in games, which have rules but also require a tacit understanding of them, and all this provides a grammar for technology use and meaning«. Übertragen auf den Kontext digitaler Wargames fehlt eine solche Grammatik auffällig – ein Punkt, der sich auch schon bei Philipp Sabin (2014, 26–27) findet, als er feststellte, dass digitale Spiele funktionieren, ohne dass ihre Benutzer:innen notwendigerweise verstehen, »was ›unter der Haube‹ vor sich geht« (Übersetzung durch die Verfass.), während manuelle Spiele in ihren Regelsystemen offener und expliziter sind.

Als »exercise in human interaction« (Perla 1990, 164) zeichnen sich Wargames durch die Präsenz des Menschen aus: »human players, immersed in scenarios, bounded by rules, and motivated by consequence-based outcomes« (Lin-Greenberg et al. 2022, 85). Wenn Wargames nun zu einer reinen Simulation werden, ist dies, wie skizziert zutiefst problematisch. Im Kern soll Entscheidungsfindung unter Unsicherheit auch ethische Unbestimmtheiten ans Licht bringen,

anstatt sie zu beseitigen. Wie Rubel (2006) argumentiert, zeichnen sich Wargames grundsätzlich durch einen Mangel an struktureller Solidität aus – dies gilt sowohl für das Instrument selbst als auch für das von ihm generierte Wissen. Dementsprechend lässt sich eine kritische konzeptionelle Unterscheidung treffen: Wargames generieren Erkenntnisse, nicht Beweise. Die epistemologische Legitimität des von ihnen produzierten Wissens – die angesichts der Risiken von LLMs kritisch hinterfragt werden muss – liegt in seiner Bedingtheit. Sie wird nicht nur durch das Szenario oder die Eigenschaften der Spieler:innen geprägt, sondern durch die gelebte Erfahrung der Entscheidungsfindung selbst. Emery (2021a) etwa betont, dass solche Entscheidungen immer auch von moralischer Handlungsfähigkeit durchdrungen sind. Die Erfahrung, die moralische Last der eigenen Entscheidungen zu tragen, wie Emerys (2021a) Rekonstruktion von Wargaming im Kalten Krieg veranschaulicht, erscheint insofern nur in qualitativen Formaten möglich, die Raum für affektives Engagement lassen. Diese schaffen eine immersive Arena, in der ethische Reflexion und moralisches Urteilsvermögen entstehen können. Während einige Autoren (vgl. z.B. Dillion et al. 2023), meinen, erste Anzeichen dafür identifiziert zu haben, dass LLMs menschenähnliches moralisches Denken zeigen, ist es plausibler anzunehmen, dass KI-Technologien derzeit nicht den Status moralischer Subjekte besitzen (vgl. z.B. Moos 2021, 243), nicht zuletzt, weil Computersystemen menschliche Empathie fehle (Emery 2021a). Diese gilt jedoch, wie Emery (2021a, 18) ausführt – und z.B. auch Veziridis et al. (2017) betonen – für ethische Entscheidungen unerlässlich und unverzichtbar für die Wahrung der Grundprinzipien der Kriegführung. Daraus wird dann auch die Schlussfolgerung gezogen, dass »[h]ow players respond to being forced to operate in morally gray areas is well beyond the ability of modern AI and psychological models« (Seater 2021, 135).

⇒ 5 Diskussion

In der militärischen Gemeinschaft, in der Wargaming als zentrales pädagogisches und Trainingsinstrument gilt, gewinnt die Idee, LLMs mit dem Entwurf von Spielszenarien zu beauftragen, durch sie Spieler:innen zu simulieren oder als Assistenten zu etablieren, die in der Lage sind, große Mengen komplexer Daten mit quasi-qualitativer Präzision und Flexibilität zu analysieren, zunehmend an Bedeutung. Diese Entwicklung wird vor allem durch die Suche nach Instrumenten vorangetrieben, die auf Konfliktszenarien mit wachsenden Abhängigkeiten und Unsicherheiten reagieren können. Sie ist jedoch mit dem noch weit-

gehend unerforschten Phänomen konfrontiert, wie sich solche Instrumente auf Praxis und Wahrnehmung auswirken. Aufbauend auf einer Darstellung der erkenntnistheoretischen Grundlagen pädagogisch motivierten Wargamings – wobei die scheinbar paradoxe Beziehung zwischen Szenario und Immersion sowie die dualen Prämissen qualitativer und quantitativer Wargames illustriert wurden – zeigte sich, dass die mit dem Einsatz von LLMs verbundenen Versprechen weitgehend auf ihrer Fähigkeit beruhen, beispiellose Datenmengen zu verarbeiten, die während ihrer Trainingsphase verwendet werden, und gleichzeitig bemerkenswerte analytische Fähigkeiten in Bezug auf ihre Spielzüge in analytischen Spielen zu zeigen. Besonders aufschlussreich erscheint das rasante Tempo der technologischen Entwicklung und die Notwendigkeit, neue Technologien für strategische und taktische Vorteile in einem zunehmend von Komplexität geprägten militärischen Kontext zu nutzen. Dies hat knapp ein Jahrzehnt nach der Renaissance des Wargamings erneut zu einer erkennbaren Verschiebung hin zur Automatisierung des Wargamings geführt – möglicherweise auf Kosten des Primats der Subjektivität. Vor diesem Hintergrund wurden die problematischen Auswirkungen und Risiken der Integration von LLMs in Wargaming aufgezeigt, die von der Verantwortungsdiffusion über die Intransparenz von Entscheidungsprozessen und dem De-Skilling bis hin zur Frage moralischer Agency reichten. Gezeigt werden konnte, weshalb allein aufgrund der technischen Einschränkungen von LLMs und der ihnen eigenen »Spiel«-Logik bisweilen gefordert wird, sie nur in Kontexten einzusetzen, in denen »getting wrong answers doesn't matter«, d.h. dort, wo sie mithin als tolerabel oder gar belanglos angesehen werden können (Chapman 2023, 77). Vor dem Hintergrund des Anspruchs von Wargames sind überdies Einwände formuliert worden, die eine Belanglosigkeit des Kontext Wargaming wiederum verneinen lassen.

Darüber hinaus ist das Potenzial von LLMs die epistemischen Grundlagen des Wargamings selbst in Frage zu stellen in den Fokus gerückt. Der Fall digitalisierter Wargames übernimmt dabei eine illustrative Funktion für die ethische Debatte über die Übertragung technischer Handlungsfähigkeit auf menschliche Handlungssysteme. An ihrem Beispiel illustriert sich überdies, dass menschliche Handlungsfähigkeit nicht zuletzt auf Entscheidungskompetenzen basiert, die sowohl aus analytischen als auch aus emotionalen Urteilen hervorgehen. Diese affektive Dimension des Be- und Urteilens gilt als spezifischer Freiheitsgrad des Spiels, dessen Erleben – und das Trainieren der Kontingenzen von Entscheidungen durch antizipatorische – sowohl synthetische als auch immersive – Erfahrungsformen die zentrale Begründung von

Wargames in militärischen Kontexten bildet. In diesem Sinne bleibt der »Human in the Loop« sein konstitutives Element. In aktuellen wissenschaftlichen Debatten zur Ethik von KI wächst gegenwärtig jedoch die Sorge über das Verschwinden des Paradigmas des »Human in the Loop« oder sogar des »Human on the Loop« (vgl. Jenkins et al. 2025). Die Sinnhaftigkeit dieser Konzepte wird aufgrund ihrer Unbestimmtheit und der Beobachtung, dass automatisierte Systeme mittlerweile mit einer Geschwindigkeit arbeiten, die mit menschlicher Kontrolle nicht mehr vereinbar ist, sogar zunehmend in Frage gestellt (Jenkins et al. 2025, 11; vgl. Emery 2021b). Vor diesem Hintergrund scheint die Diskussion über LLMs als Technologie im breiteren KI-Bereich im Wargaming geradezu die Relevanz des menschlichen Faktors und damit des »Human in the Loop« in seiner grundlegendsten Konzeption zu unterstreichen: als »giving the right value to the knowledge producers«, wie Fabio Massimo Zanzotto (2019, 244) es formuliert – obwohl er dies vor allem im Hinblick auf den wirtschaftlichen Wert des von Menschen bereitgestellten Wissens für KI-Systeme betont. Gerade weil die Frage, wer welches Wissen mit welchen Konsequenzen produziert, von so grundlegender Bedeutung ist, ist es unerlässlich, das Konzept des »Human in the Loop« in die Bedingungen und Konsequenzen der Mensch-Maschine-Interaktion einzuordnen (vgl. Emery 2021b), wie es etwa auch Sicart (2020, 22) für die Ethik des Play im Informationszeitalter formuliert: »The homo poieticus is a steward of the values and informational integrity of the environment they inhabit«.

Führt man diese Blickrichtungen nun zurück auf die eingangs aufgeworfene Frage nach dem Realen und Baudrillards Diagnose, dass Simulationen schließlich realer als das Reale werden, erhält diese im Kontext automatisierter Entscheidungshilfen eine neue Dringlichkeit. Wenn man nämlich die hier diskutierten Risiken der Integration von LLMs in das Wargaming – von algorithmischer Opazität und halluzinativer Sinnstiftung über Verantwortungsdiffusion bis hin zur potenziellen Erosion menschlicher Urteilsfähigkeit – in Baudrillards Theoriehorizont zurückspiegelt, lässt sich die gegenwärtige Entwicklung als eine weitere Verschiebung hin zu einem fortgeschritteneren Stadium des Simulakrums deuten: LLM-gestützte Wargames produzieren nicht lediglich ein neues Werkzeug militärischer Analyse oder Ausbildung. Vielmehr tragen sie dazu bei, eine zusätzliche Ebene des Hyperrealen zu erzeugen: in dieser verdoppelt die Simulation nicht nur die Realität des Krieges, sondern sie riskiert die eigenen epistemologischen und normativen Koordinaten aufzuheben. Dies ist nicht nur dadurch impliziert, dass sich das Militärische zunehmend an einer technisch generierten Realität orientiert, deren Referenz Datenmodelle sind. Je mehr LLMs dem

Wargaming seine kontingenten, relationalen und emotional erfahrbaren Momente entziehen, desto stärker bewegt sich das Feld in Richtung einer rein synthetischen Realität, in der Handlungen nicht mehr nur vorweggenommen, sondern in der sie auch präfiguriert werden. Damit lässt sich auch ein weiterer Kreis schließen, der mit dem Gedicht von Noomi Stone eröffnet wurde. Obwohl mit ihrer Arbeit und diesem Beitrag unterschiedliche Formen des Wargaming untersucht wurden, ist mit der Perspektive des Vorliegenden dennoch eine vergleichbare Einrede vollzogen, die Stone 2022 formuliert hat:

While the military bases the »game« on the »real«, certain forms of the lived, embodied, affective, and political real are deemed superfluous or divisive, exterior to the tactical goals of the game. This delicate tension creates a porous vertigo: Which forms and degrees of knowledge and emotion will be granted realness? (Stone 2022, 118)

Doch was ist nun aus jener Einrede für diese Arbeit zu schlussfolgern? Für die Beantwortung dieser Frage erweist sich John J. Hortons (2023, 2) Hinweis als nützlich: Obwohl alle Modelle – LLMs eingeschlossen – naturgemäß fehlerhaft sind, schließt dies nicht aus, dass sie uns dabei helfen, über die Art unserer Fragen und deren Beantwortung nachzudenken. Wenn Digitalisierung und Automatisierung im militärischen Umfeld zur Norm werden, stellt sich die drängende empirische Frage, wie sich der Einsatz von LLMs – insbesondere in unterstützenden Rollen – auf tatsächliche Entscheidungsprozesse auswirkt und wie Spieler:innen die (Konsequenzen dieser) Interaktivität erleben. Es gilt dann, mit Sicart (2020) (resp. Floridi) gesprochen, die re-ontologisierenden Fähigkeiten (2020) der KI im Spiel und durch das Spiel erfahrbar werden zu lassen. Schließlich sind seine Spieler:innen gezwungen Entscheidungen zu treffen, die auf konsequenzbasierte Ergebnisse reagieren oder zu solchen führen. Fragen ließe sich dann nicht nur ob, sondern wie eine Zunahme technologischer Handlungsfähigkeit mit einem Verlust an professionellem Urteilsvermögen und Handlungsfähigkeit einhergeht. Durch eine solche Befasstheit mit der Idee der Automatisierung verschöbe sich die Funktion von LLMs in Wargames weg von einer das Objektivitätsideal bedienenden Logik zur Eliminierung menschlicher Fehler hin zu ihrer Neukonzeptualisierung als »real world stimuli« (Lin-Greenberg et al. 2022, 88), d.h. als Kontextvariable im Wargaming. Eine solche Untersuchung wäre von Natur aus ethisch und politisch, denn

in thinking of play as simulational [...] simulations and virtual realities are real in their technical apparatus and their effects on our bodies. Play and playful technologies might help us rethink artificial realities of culture and communication – and the reality of the artifice (Giddings 2016, 155).

Im Kontext von Wargaming im Zeitalter der KI ist es daher unerlässlich, das Spiel – oder vielmehr die Simulation selbst – zu befragen, indem man seine und ihre Regeln testet, verbiegt, ersetzt oder entfernt, um die Erkenntnisse und Effekte zu erforschen, die dies hervorbringen kann (Hirst 2024, 152). Solch eine kritische Reflexion sensibilisiert uns für die Frage, ob es überhaupt richtig – oder gut – war, dieses spezielle Spiel zu spielen, denn wie Hirst (2024, 152, H.i.O.) es ausdrückt: »more important than *what* we play is *how* we play.«

⇒ Literaturverzeichnis

Aguirre, Anthony/Javorsky, Emilia/Tegmark, Max (2023): ›Artificial Escalation‹: Imagining the future of nuclear risk, in: Bulletin of the Atomic Scientists, Download unter: <https://thebulletin.org/2023/07/artificial-escalation-imagining-the-future-of-nuclear-risk/> (Zugriff am 05. September 2025).

Agnew, William/Bergman, A. Stevie/Chien, Jennifer/Díaz, Mark/El-Sayed, Seliem/Pittman, Jaylen/Mohamed, Shakir/McKee, Kevin R. (2024): The Illusion of Artificial Inclusion, in Mueller, Florian Floyd/Kyburtz, Penny/Williamson, Julie R./Sas, Corina/Wilson, Max L.M./Toups Dugas, Phoebe/Shklovski, Irina (Hg.): CHI'24: Proceedings of the 2024 CHI Conference on Human Factors in Computing System, New Yor: Association for Computing, Article 286: 1–12. <https://doi.org/10.1145/3613904.3642703>.

Barzashka, Ivanka (2019): Wargaming: how to turn vogue into science, in: Bulletin of the Atomic Scientist. March 15, 2019, Download unter: <https://thebulletin.org/2019/03/wargaming-how-to-turn-vogue-into-science/> (Zugriff am 05. September 2025).

Baudrillard, Jean (2010): Simulacra and Simulation. Translated by Sheila Faria Glaser. Ann Arbor: The University of Michigan Press. <https://doi.org/10.3998/mpub.9904>.

Baudrillard, Jean (1991): The Gulf War Did Not Take Place. Translated and with an introduction by Paul Patton. Bloomington/Indianapolis: Indiana University Press.

Black, Scotty/Darken, Christian (2024): Scaling Artificial Intelligence for Digital Wargaming in Support of Decision-Making, STO-MP-MSG-207, Download unter: <https://arxiv.org/pdf/2402.06075> (Zugriff am 05. September 2025).

CGSC, Unethical Professional Wargaming (2021): make your war-game say what you want it to say. Connections US 2021 Wargaming Conference. US Command and General Staff College and The CGSC Foundation. Final Report, 13th September 2021, Download unter: <https://paxsims.wordpress.com/wp-content/uploads/2021/09/unethical-professional-wargaming-final-report-20210913-v2.pdf> (Zugriff am 05. September 2025).

Chapman, David (2023): Better without AI. New York: Cloud Pattern Press.

Chen, Yuwei/Chu, Shiyong (2024): Large Language Models in Wargaming: Methodology, Application, and Robustness. *EEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. <https://doi.org/10.1109/CVPRW63382.2024.00295>.

Coeckelbergh, Mark (2022): The Grammars of AI: Towards a Structuralist and Transcendental Hermeneutics of Digital Technologies, in: *Technology and Language*, 3(2), 148–161. <https://doi.org/10.48417/technolang.2022.02.09>.

Collins, Harry (2018): *Artificial Intelligence. Against Humanity's Surrender to Computers*. Cambridge: Polity Press.

Curry, John/Engle, Chris/Perla, Peter (2020): *The Matrix Games Handbook: Professional Applications from Education to Analysis and Wargaming*. Morrisville, NC: History of Wargaming Project.

Davis, Paul K./Bracken, Paul (2025): Artificial Intelligence for Wargaming and Modelling, in: *Journal of Defense Modelling and Simulation: Applications, Methodology, Technology* 22(1), 25–40. <https://doi.org/10.1177/15485129211073126>.

Dillion, Danica/Tandon, Niket/Gu, Yuling/Gray, Kurt (2023): Can AI language models replace human participants? in: *Trends in Cognitive Sciences* 27(7), 97–600. <https://doi.org/10.1016/j.tics.2023.04.008>.

Doktrinzentrum der Bundeswehr (2024): *Wargaming Handbuch Bundeswehr*, Berlin, Download unter: <https://www.bundeswehr.de/resource/blob/5872302/f33665313cbbf07c098ed7e3769530ae/handbuch-wargaming-bundeswehr-data.pdf> (Zugriff am 05. September 2025).

Downes-Martin, Stephen (2013): The Diabolus in Machina of War Gaming, in: *Naval War College Review*, 66(3), 67–80. <https://www.jstor.org/stable/26397397?seq=1>.

DWDS (2025a): Simulation, Download unter: <https://www.dwds.de/wb/Simulation#etymwb-1> (Zugriff am 04. September 2025).

DWDS (2025b): Spiel, Download unter: <https://www.dwds.de/wb/Spiel#etymwb-1> (Zugriff am 04. September 2025).

Emery, John R. (2021a): Moral Choices without Moral Language. 1950s Political-Military Wargaming at the RAND Corporation, in: *Texas National Security Review* 4(4).

Emery, John R. (2021b): Algorithms, AI, and Ethics of War, in: *Peace Review*, 33(2), 205–212. <https://doi.org/10.1080/10402659.2021.1998749>.

Esposito, Elena (2024): *Kommunikation mit unverständlichen Maschinen*. Wien; Salzburg: Residenz-Verlag.

FAIR (2022): Human-level play in the Game of Diplomacy by combining language models with strategic reasoning, in: *Science* 378(6624), 1067–1074. <https://doi.org/10.1126/science.ade9097>.

Felder, Ekkehard/Kückelhaus, Marcel (2025): Das definierende Sprachmodell (LLM): Anthropomorphisierung in der Mensch-Maschine-Interaktion, in: *Zeitschrift für Literaturwissenschaft und Linguistik* 55, 431–448. <https://doi.org/10.1007/s41244-025-00380-7>.

Fight Club International (2025): Persian Predicament. A Matrix Game on a potential Iranian future. Post Game Report, Download unter: <https://www.fightclubinternational.org/persian-predicament-iran-matrix-game> (Zugriff am 04. September 2025=).

Gallotta, Robert/Todd, Graham/Zammit, Marvin/Earle, Sam/Liapis, Antonios/Togelius, Julian/Yannakakis, Georgios N. (2024): Large Language Models and Games: A Survey and Roadmap, accepted for publication at the *IEEE Transactions on Games*. <https://doi.org/10.48550/arXiv.2402.18659>.

Giddings, Seth (2016): *Gameworlds: Virtual Media and Children's Everyday Play*. New York: Bloomsbury Academic.

Goodman, James/Risi, Sebastian/Lucas, Simon (2020): AI and War-gaming. <https://doi.org/10.48550/arXiv.2009.08922>.

Grossmann, Igor/Feinberg, Matthew/Parker, Dawn C./Christakis, Nicholas A./Tetlock, Philip E./Cunningham, William A. (2023): AI and the Transformation of Social Science Research. Careful bias management and data fidelity are key, in: *Science*, 380(6650), 1108–1109. <https://doi.org/10.1126/science.adi1778>.

Hagman, Michelle/Sacco, Kathrine (2017): Kindred Tools. An Interview with Nomi Stone. Supplemental, in: *Fieldsights*, 26. April 2017, Download unter: <https://www.culanth.org/fieldsights/kindred-tools-an-interview-with-nomi-stone> (Zugriff am 03.09.2025).

Harper, Douglas (2025a): Etymonline: Game, Download unter: <https://www.etymonline.com/search?q=game> (Zugriff am 04. September 2025).

Harper, Douglas (2025b): Etymonline: Play, Download unter: <https://www.etymonline.com/search?q=play> (Zugriff am 04. September 2025).

Helberger, Natali/Araujo, Theo/de Vreese, Claes (2020): Who is the fairest of them all? Public attitudes and expectations regarding automated decision-making, in: *Computer Law & Security Review* 39, 105456. <https://doi.org/10.1016/j.clsr.2020.105456>.

Hirst, Aggie (2024): *Politics of play. Wargaming with the US Military*. New York: Oxford University Press, 2024. <https://doi.org/10.1093/oso/9780197629192.001.0001>.

Hoffman, Wyatt/Kim, Heu Millie (2023): Reducing the Risks of Artificial Intelligence for Military Decision Advantage. Policy Brief. CSET, Download unter: <https://cset.georgetown.edu/publication/reducing-the-risks-of-artificial-intelligence-for-military-decision-advantage/> (Zugriff am 03. September 2025).

Hogan, Daniel P./Brennen, Andrea (2024): Open-Ended Wargames with Large Language Models. <https://doi.org/10.48550/arXiv.2404.11446>.

Horton, John J. (2023): Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus? (Working Paper 31122), National Bureau of Economic Research, Download unter: https://www.nber.org/system/files/working_papers/w31122/w31122.pdf (Zugriff am 03.09.2025).

Hovd, Sigurd (2023): Tools of war and virtue – Institutional structures as a source of ethical deskilling, in *frontiers*. *Frontiers in Big Data* 5, 1019293. <https://doi.org/10.3389/fdata.2022.1019293>

Huizinga, Johan (2022): *Homo Ludens. Vom Ursprung der Kultur im Spiel*. Reinbek bei Hamburg: Rowohlt.

Iser, Wolfgang (1976), *Der Akt des Lesens. Theorie Ästhetischer Wirkung*. Stuttgart: UTB-Verlag.

Jahanbakhsh, Maziar/Kochaki, Hassan/Shamami, Nader (2023): Effective factors in the design and execution of the war game, in: *Iranian Journal of Wargaming* 6(12). <https://doi.org/10.22034/qjmst.2023.544038.1627>.

Jenkins, Ryan/Sullins, John P./Kalu, Obinna/Kamath, Aisha/Phumjam, Kritтана (2025): Recent Insights in Responsible AI Development and Deployment in National Defense: A Review of Literature 2022–2024,

in: *Journal of Military Ethics* 24(19), 63–85. <https://doi.org/10.1080/15027570.2025.2483058>.

Kaminski, Andreas/Gramelsberger, Gabriele/Scheer, Dirk (2023): Introduction. Modeling for policy and technology assessment: Challenges from Computer-based simulations and artificial intelligence, in: *TATuP. Journal for Technology Assessment in Theory and Practice* 31(1), 11–17.

Lamparth, Max/Corso, Anthony/Ganz, Jacob/Mastro, Oriana Skylar/Schneider, Acquelyn/Trunkunas, Harald (2024): Human vs. Maschine: Behavioral Differences between Expert Humans and Language Models in Wargame Simulations, in: Das, Sanmay/Green, Brian Patrick/Vasheney, Kush/Ganapini, Marianna/Renda, Andrea (Hg.): *Proceedings of the Seventh AAAI/ACM Conference on AI, Ethics, and Society (AIES2024)*, Washington: The AAAI Press, 807–817. <https://doi.org/10.48550/arXiv.2403.03407> (Zugriff am 03. September 2025).

Licato, John (2022): War-Gaming Needs Argument-justified AI More Than Explainable AI, in: Sen, Atriya (Hg.): *Proceedings of 2022 Advances on Societal Digital Transformation (DIGITAL) Special Track on Explainable AI in Societal Games (XAISG)*, International Academy, Research, and Industry Association (IARIA).

Lin, Zichao/Guan, Shuyan/Zhang, Wending/Zhang, Huiyan/Li, Yugang/Zhang, Huaping (2024): Towards trustworthy LLMs: a review on debiasing and dehallucinating in large language models, in: *Artificial Intelligence Review* 57, 243. <https://doi.org/10.1007/s10462-024-10896-y>.

Lin-Greenberg, Erik/Paul, Reid B.C./Schneider, Jacquelyn G. (2022): Wargaming for International Relations research, in: *European Journal of International Relations* 28(1), 83–109.

Liu, Furong/Wei, Tingting/Zhang, Zhiheng/Chen, Jiaying/Lu, Lina (2024): Survey of Opponent Modelling: from Game AI to Combat Deduction, in: *IEEE (Hg.): 36th Chinese Control and Decision Conference (CCDC)*. <https://doi.org/10.1109/CCDC62350.2024.10588019>.

Luhmann, Niklas (2017): *Die Realität der Massenmedien*. Wiesbaden: Springer VS.

McDermott, Rose (2002): Experimental Methodology in Political Science, in *Political Analysis* 10(4), 325–342. <https://doi.org/10.1093/pan/10.4.325>

McKinnon, Douglas J./Zhao, Ying/Kendall, Walter Anthony/Schwamm, Riqui (2022): Leverage AI to Learn, Optimize, and Wargame (LAILOW) for Strategic Laydown and Dispersal (SLD) of the USN Operating Forces. Reports and Technical Reports. Monterey: Naval Postgraduate School, Download unter: <https://core.ac.uk/download/565849501.pdf> (Zugriff am 03.09.2025).

Mitsea, Eleni/Drigas, Athanasios/Skianis, Charalabos (2025): A Systematic Review of Serious Games in the Era of Artificial Intelligence, Immersive Technologies, the Metaverse, and Neurotechnologies: Transformation Through Meta-Skills Training, in: *Electronics* 14, 649. <https://doi.org/10.3390/electronics14040649>.

Moos, Thorsten (2021): Digitaler Animismus. Theologische Bemerkungen zu einer Ethik der Digitalisierung. In: Held, Benjamin; von Oorscot, Frederike (Hg.): *Digitalisierung: Neue Technik, neue Ethik?*, Heidelberg: heiBOOKS, 235–256. <https://doi.org/10.11588/heibooks.945.c12684>

Mouat, Tom (2022): The Use and Misuse of Wargames, in: *Scandinavian Journal of Military Studies* 5(1), 209–220. <https://doi.org/10.31374/sjms.121>.

Nitzl, Christian/Landsiedel, Jan/Borghoff, Uwe M. (2023): Lernen aus den Erfahrungen von militärischen Wargames, in: *Wirtschaftsinformatik & Management* 15(5), 347–353. <https://doi.org/10.1365/s35764-023-00494-y>.

Perla, Peter P. (2016): Operations Research, Systems Analysis, and Wargaming: Riding the Cycle of Research. In: Harrigan, Pat/Kirschenbaum, Matthw G. (Hg.): *Zones of Control. Perspectives on Wargaming*, Cambridge; London: The MIT Press, 159–182.

Perla, Peter (1990): *The Art of Wargaming: A Guide for Professionals and Hobbyists*. Annapolis: Naval Institute Press.

Pugliese, Joseph (2017): How Drones Are Gamifying War in America's Casino Capital. USAPP–American Politics and Policy, Download unter: <https://blogs.lse.ac.uk/usappblog/2017/09/15/how-drones-are-gamifying-war-in-americas-casino-capital/> (Zugriff am 03.09.2025).

Pugliese, Joseph (2016): Drone Casino Mimesis: Telewarfare and Civil Militarization, in: *Journal of Sociology*, 52(3).

Rivera, Juan-Pablo/Mukobi, Gabriel/Reuel, Anka/Lamparth, Max/Smith, Chandler/Schneider, Jacquelyn (2024): Escalation Risks from Language Models in Military and Diplomatic Decision-Making, in:

Association for Computing Machinery (Hg.): ACM FAccT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency, New York: Association for Computing Machinery, 836–898. <https://doi.org/10.1145/3630106.3658942>.

Rubel, Robert C. (2006): The Epistemology of War Gaming, in: *Naval War College Review* 59(2), 108–128, Download unter: <https://www.jstor.org/stable/26396724> (Zugriff am 03.09.2025).

Sabin, Philipp (2014): *Simulating War. Studying Conflict Through Simulation Games*. London; New York: continuumbooks. <https://doi.org/10.5040/9781474211239>.

Schreiber, Gerhard (2025): Beyond Code – The Crucial Role of Responsibility in Technology Ethics. Teilkapitel in Söllner, Matthias/Arnold, Thomas/Benlian, Alexander/Bretschneider, Ulrich/Knight, Caroline/Ohly, Sandra/Rudkowski, Lena/Schreiber, Gerhard/Wendt, Domenik: *ChatGPT and Beyond: Exploring the Responsible Use of Generative AI in the Workplace*, in: *Business & Information Systems Engineering* 67, 289–303.

Seater, Robert (2021): Can AI save Us from Unethical Wargaming? In: CGSC, *Unethical Professional Wargaming*. »make your wargame say what you want it to say«, Connections US 2021 Wargaming Conference. US Command and General Staff College and The CGSC Foundation. Final Report, 13th September 2021, 125–138, Download unter: <https://paxsims.wordpress.com/wp-content/uploads/2021/09/unethical-professional-wargaming-final-report-20210913-v2.pdf> (Zugriff am 03. September 2025).

Shrivastava, Aryan (2024): Response Inconsistency of Large Language Models in High-Stakes Military Decision Making, Download unter: https://bpb-us-w2.wpmucdn.com/voices.uchicago.edu/dist/9/3887/files/2024/10/SRF24_FINAL_Shrivastava_Aryan.pdf (Zugriff am 03. September 2025).

Sicart, Miguel (2020): *Homo Ludens Reloaded: The Ethics of Play in the Information Age*, in: Groen, Maïke/Kiel, Nina/Tillmann, Angela/Weßel, André (Hg.): *Games and Ethics. Theoretical and Empirical Approaches to Ethical Questions in Digital Game Cultures*, Wiesbaden: VS Verlag, 13–27.

Sicart, Miguel (2016): *We the Soldiers: Play Complicity and Ethical Gameplay in Call of Duty: Modern Warfare*. In: Harrigan, Pat/Kirschenbaum, Matthew G. (Hg.): *Zones of Control. Perspectives on Wargaming*. Cambridge; London: The MIT Press, 309–318.

Sicart, Miguel (2014). *Play Matters*. Cambridge; London: The MIT Press.

Silver, David/Huang, Aja/Maddison, Chris J./Guez, Arthur/Sifre, Laurent/van den Driessche, George/Schrittwieser, Julian/Antonoglou, Ioannis/Panneershelvam, Veda/Lanctot, Marc/Dieleman, Sander/Grewe, Dominik/Nham, John/Kalchbrenner, Nal/Sutskever, Ilya/Lillicrap, Timothy/Leach, Madeleine/Kavukcuoglu, Koray/Grapel, Thore/Hassabis, Demis (2016): Mastering the game of Go with deep neural networks and tree search, in: *nature* 529, 484–489. <https://doi.org/10.1038/nature16961>.

Stone, Nomi (2022): *Pinelandia: An Anthropology and Field Poetics of War and Empire*. Oakland: University of California Press. <https://doi.org/10.2307/j.ctv2vr8tz5>.

Stone, Nomi (2019): *Kill Class*. Noth Adams: Tupelo Press.

Thompson, Nicholas (2024): *Strategic Gaming with AI Copilots: Balancing Assistance and Autonomy*. Dissertation, University of Bath, 2024.

Vallor, Shannon (2015): Moral Deskillling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character, in *Philosophy & Technology* 28(1), 10–124. <https://doi.org/10.1007/s13347-014-0156-9>.

Veiziridis, Savvas/Karampelas, Panagotis/Lekea, Ioanna (2017): Learn by Playing. A serious war game simulation for teaching military ethics, in *IEEE (Hg.): IEEE Global Engineering Education Conference (EDUCON)*, Athens, Greece, 920–925.

Wacquant, Loïc (2009): The Body, the Ghetto and the Penal State, in: *Qualitative Sociology* 32, 101–129. <https://doi.org/10.1007/s11133-008-9112-2>.

Walther, Bo Kampmann (2003): Playing and Gaming. Reflections and Classifications, in: »Games Studies«, *The International Journal of Computer Game Research*. 3(1), Download unter: <https://www.game-studies.org/0301/walther/> (Zugriff am 03. September 2025).

Weidinger, Laura/Uesato, Jonathan/Rauh, Maribeth/Griffin, Connor/Huang, Po-Sen/Mellor, John/Glase, Amelia/Cheng, Myra/Balle, Borja/Kasirzadeh, Atoosa/Biles, Courtney/Brown, Sasha/Kenton, Zac/Hawkins, Will/Stepleton, Tom/Birhane, Abeba/Hendricks, Lisa Anne/Rimell, Laura/Isaac, William/Haas, Julia/Legassick, Sean/Irving, Geoffrey/Gabriel, Iason (2022): Taxonomy of Risks posed by Language

Models, in: FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 214–229. <https://doi.org/10.1145/3531146.3533088>.

Weller, Dominic/Meltschak, Max/Schwindling, Dominik (2024): Leveraging Large Language Models for Enhanced Wargaming in Multi-Domain Operations. NATO STO-MP-MSG-217, Download unter: <https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-MSG-217/MP-MSG-217-16.pdf> (Zugriff am 03.09.2025).

Wilden, Adam/Nasim, Mehwish/Williams, Peter/Legrand, Tim/Turnbull, Benjamin Peter/Williams, Patricia A.H. (2023): On Benchmarking and Validation in Wargames, in: Andreatos, Antonios/Dougligeris, Christos (Hg.): Proceedings of the 22nd European Conference on Cyber Warfare and Security, ECCWS, 533–543. <https://doi.org/10.34190/ec-cws.22.1.1132>.

Yu, Sheng/Zhu, Wei/Wang, Yong (2023): Research on Wargame Decision-Making Method Based on Multi-Agent Deep Deterministic Policy Gradient, in: Applied Sciences 13(7), 4569. <https://doi.org/10.3390/app13074569>.

Zanzotto, Fabio Massimo (2019): Human-in-the-loop. Artificial Intelligence, in Journal of Artificial Intelligence Research 64, 243–252. <https://doi.org/10.1613/jair.1.11345>.

Zhang, Yadong/Mao, Shaoguang/Ge, Tao/Wang, Xun/de Wynter, Adrian/Xia, Yan/Wu, Wenshan/Song, Ting/Lan, Man/Wie, Furu (2024): LLM as a Mastermind: A Survey of Strategic Reasoning with Large Language Models, preprint. <https://doi.org/10.48550/arXiv.2404.01230>.

Zuber, Kenneth Wade (2024): The Experiences of Wargame Participants: A Phenomenological Study of How Participants ascribe Meaning to their Experience. Dissertation, University of Louisiana Monroe. ProQuest 31639463.

Zwarts, Natalia (2024): Wargaming as a discipline that constantly develops. How to ensure that new formats are improving the practice?, in: Revista de la Escuela Superior de Guerra Naval 21(1), 117–126, Download unter: <https://revista.esup.edu.pe/RESUP/article/view/194> (Zugriff am 03.09.2025).

Zitationsvorschlag:

Kühne, Sylvia (2026): Playful technologies? Über Anspruch und Risiko von Künstlicher Intelligenz im Wargaming (Ethik und Gesellschaft 1/2026: Kein Spiel. Wargaming und Serious Gaming im Zeitalter der KI). Download unter: <https://dx.doi.org/10.18156/eug-1-2026-art-6> (Zugriff am [Datum]).



ethikundgesellschaft
ökumenische zeitschrift für sozialetik

1/2026: Kein Spiel. Wargaming und Serious Gaming im Zeitalter der KI

Gerhard Schreiber
Kein Spiel. *Wargaming* und *Serious Gaming* im Zeitalter der KI. Zur Einleitung

Lukas Johrendt und Kathrin Bruder
(K)eine Spiel-Moral? Ethik diesseits und jenseits des Kriegsspiels

Benedikt Bussmann
Spielerisch in den Abgrund blicken

Isabelle Fries
Ein Spielverderber namens Ernst
Metaethische Reflexionen zu Wargames und Serious Games

Lukas Ohly
Können KI-Trainingsspiele Kriege humanisieren?

Sylvia Kühne
Playful technologies?
Über Anspruch und Risiko von Künstlicher Intelligenz im Wargaming

Marie-Christin Barleben
Peacegaming. Oder wie wir spielend Frieden lernen

Max Tretter
Digital Sovereignty and Video Games