WHEN ROBOTS MET LANGUAGE (MODELS): AN EXPLORATION OF SCIENCE AND STRATEGY IN THE *VISION-LANGUAGE-ACTION* MODELS SPACE

Documents de travail GREDEG
GREDEG Working Papers Series

Taekyun Kim
Gabriela Fuentes
Simone Vannuccini

GREDEG WP No. 2025-33
https://ideas.repec.org/s/gre/wpaper.html

When Robots Met Language (Models): An Exploration of Science and Strategy in the *vision-language-action* Models Space*

Taekyun Kim[1], Gabriela Fuentes[2], and Simone Vannuccini[3]

[1]Chungnam National University
[2,3]GREDEG, CNRS, Université Côte d'Azur

July 31, 2025

Abstract

We study the emerging technology of vision-language-action models (VLAMs), a potential breakthrough for the robotics industry. VLAMs connect artificial intelligence's large language models with robots: they are a software innovation that increases flexibility and expand capabilities of robotic hardware. As the technology is yet in its infancy, we focus our analysis on three directions. First, we study the technological trajectory of VLAMs in scientific publications using scientometric techniques. Second, in order to assess market transformations associated with VLAMs and potential migration in the locus of value generation in robotics, we collect information and offer insights on the actors developing the technology 'in the wild'. Third, we discuss the strategies put in place by a leading actor in the field to gatekeep the competitive landscape. We then generalise and rationalise our findings through established theoretical frameworks. The paper is the first to single out the forces at work in the development of an innovation poised to rejuvenate the robotics industry, one of the backbones of contemporary economies.

Keywords: robotics; artificial intelligence; vision-language-action models; VLA; embodied AI

JEL Codes: L10; L86; 030.

*Corresponding author: Simone Vannuccini simone.vannuccini@univ-cotedazur.fr. We are grateful to Ekaterina Prytkova, Ludovic Dibiaggio and Michele Pezzoni for the help and feedback provided in different occasions along the development of the project. We also thank the attendants to CORA Conference 2024, GREDEG PhD Days 2025 and Manchester Institute of Innovation Research (MIOIR) seminars for the valuable comments to our presentations of the research.

"The connections (between LLMs and robotics) turned out to be extremely deep, so deep in fact, that they forced us to rethink all the foundations of how we do robotics and embodied AI."

– Vincent Vanhoucke (Keynote at NVIDIA GTC 2024)

1 Introduction

Can an innovation in a given technological domain or industry induce a profound transformation — and even disruption — in another industry? Although often conflated, the fields of artificial intelligence (AI) and robotics have been evolving along largely independent trajectories. However, their paths are becoming increasingly intertwined. This has placed a new spotlight on robotics, enough to suggest that "a revolution in how robots learn" is currently unfolding.[1] The cornerstone of this potential revolution in the making are AI models tailored to robotics, and in particular *vision-language-action models* (VLAMs). VLAMs are robotic foundation models[2]: a development in the field of AI that integrates large language models (LLMs) engines into robot technology.

The importance of VLAMs for the robotics industry cannot be understated: first, conceptually, the adoption of VLAMs indicates a structural transformation in robot development, that is re-organising around the underlying logic characterising the new technology. In practice, VLAMs are "bringing problems of robotics into the semantic space". From this angle, the diffusion of VLAMs can be read through the lenses of 'redomaining', the process through which established tasks or functions start to be performed under new principles [Arthur, 2009]. In these terms, the technology exemplifies both a pattern and a case study of emerging, radical and possible disruptive innovation that is worth unpacking. Second, pragmatically, VLAMs deployment promises to enable the emergence of flexible, generalist robots with a plurality of affordances, breaking away with the industry standard of specialised equipment thanks to the removal of key bottlenecks. This is particularly the case for data: the limitation imposed by the heterogeneous data required to expand robots' capabilities is circumvented through the integration of LLMs, trained on large-scale datasets that include the whole Internet. Third, as a novel source of value creation and capture, VLAMs are a strategic lever. This means that actors' decisions and performance will be affected by the new opportunities introduced by the technology, as well as by the adaptation costs it will impose. As a result, VLAMs have the potential to reshuffle the competitive landscape of the industry: profits and market shares might migrate from hardware to software firms;

[1] https://www.newyorker.com/magazine/2024/12/02/a-revolution-in-how-robots-learn.
[2] https://sites.google.com/view/xembodimentworkshop

entry of new actors in the segment might increase; dominant firms might attempt to leverage the innovation in order to gain or reinforce control of the segment and the whole industry.

As a technology and domain of research, VLAMs belong to the growing field of embodied AI, which in general focuses on how AI software solutions are integrated in hardware devices. The growing interest in this particular trajectory of AI development is sparked by a series of factors. First, the stabilisation and commodification of the AI model market around LLMs as the dominant design reduce the uncertainty involved in exploring downstream applications, from defense to autonomous vehicles. Second, and related, the increasing demand for AI portability directs research and resources towards embodied AI. Third, and maybe even more important, the lack of profitability of software-only AI[3], despite revenue growth and the sizeable capital expenditures characterising the market, makes investments in robotic applications — not limited to VLAMs but including also to so-called humanoid robots segment[4] — rather attractive. In summary, a convergence of forces seem to suggest that AI is on the brink of a 'VLAMs moment'.

In this paper, we study the emergence of VLAMs and related models as a case of innovation at the crossroads of software and hardware domains. As the technology is yet in its infancy, we design the research as a cross-section mapping of dynamics in three complementary directions: the scientific, the market, and the strategic. In the scientific mapping, we conduct a scientometric study of the citation patterns to key VLAMs papers. We present evidence of an ongoing international race to develop the technology. Furthermore, we discuss the interesting finding that advances in VLAMs (that is, seminal scientific contributions) have been pushed in first stance by private research laboratories or private-public collaborations, with follow-on work taken up mainly by public institutions. We label this pattern a 'Bell Labs 2.0' regime, and suggest that it could be indicative of a 'return to the past' in the mode of conducting basic research, at least for certain technological fields. We also consider whether the evidence points to a case of disruption, in this instance, of the science marketplace, by private companies attempting to take over the leadership in knowledge production from Universities.

For the market mapping, we compile a dataset of companies developing or offering robot technology powered by LLMs. We track partnerships and funding to identify the sources of market dynamism in the field, and whether this domain is subject to lateral entry by large tech companies. We distinguish between VLAMs and humanoid companies to highlight

[3]https://www.bloomberg.com/news/newsletters/2024-06-19/salesforce-workday-struggle-to-make-money-from-boom-in-ai-demand

[4]https://news.crunchbase.com/robotics/ai-humanoid-robots-venture-funding-2024/

whether the two trajectories are developed jointly. We also assess the overlap between actors active in scientific and market developments, and find a rather limited intersection.

In the strategy part of the analysis, we discuss whether and how the actions taken by a key innovative and dominant actor in the VLAMs space resemble those usually taken by actors that aim at becoming technology gatekeepers, dominating the control points of an industry and maximising value capture, even through the leveraging of open-sourcing strategies. The goal of this part is also to understand whether the VLAMs segment of the robotics industry is contestable or is poised to be dominated by incumbents and large entrants that exploit their cross-industry economies of scope.

To the authors' knowledge, this is the first study exploring VLAMs, an emerging technology with the potential to rejuvenate and disrupt an established industry such as robotics. The technology is yet at the outset of its learning curve, and has a large scope for improvements, especially as scaling laws (which at the moment continue to drive advances in software AI) appear to apply to robotic foundation models as well [Sartor and Thompson, 2024]. Given the novelty of the technology, the paper is necessarily broad in scope. The value of the analysis is both explorative and illustrative. Explorative, as we shed light on the co-evolution of technology and the market in this emerging space. Illustrative, as we use the VLAM case to showcase well-known strategic behaviours in a novel context. Furthermore, as one of the focuses of our analysis is to identify whether software or hardware actors prevail in the field, we contribute to the broader discussion on the 'softwarisation' of physical industries, a transformation already detected in the automotive domain with the shift towards software-defined vehicles [Liu et al., 2022].

The paper proceeds as follows: in Section 2, we place VLAMs within the broader evolution of and recent trends in robotics. First, we overview the industry by exploiting information from a major robotics fair (*Automate Show 2024*); then, we introduce the growing field of *Embodied AI*, and highlight the emergence of our technology of interest, VLAMs, within it. In Section 3, we present our three-pronged analysis. Each subsection introduces a specific research question, the data collected, and findings. Through subsection 3.1, we examine the evolving knowledge production landscape of VLAMs using scientometric methods; in subsection 3.2, we investigate the composition and dynamics of actors driving VLAM-related innovation; and in subsection 3.3, we explore the strategic positioning and competitive behavior of a key firm in VLAM innovation, Google DeepMind, to assess how its attempts to shape the trajectory of the technology resemble ideal-typical strategy choices characterising novel markets. We rationalise our findings in Section 4, interpreting VLAM's scientific, market and strategy dynamics through the lenses of breakthrough innovation, corporate basic science, and platformisation. Finally, Section 5 summarises the key insights and implications for robotics and AI industries, and outlines future directions for both research and policy.

2 Placing VLAMs within the Evolution of Robotics

2.1 Robotics Industry Overview: Insights from Automate Show

Robotics is an established and mature industry, serving an array of application sectors, both in manufacturing (e.g., automotive, electronics, food production) and services.[5] However, in recent years the industry is undergoing a significant transformation — a 'rejuvenation' in innovation activity and market dynamism — as complementary technologies such as AI on the software side and sophisticated sensing technologies on the hardware side are improving robotic control, planning, and flexibility. To get a sense of this transformation, we capture some stylised facts from a major trade event dedicated to robotics: Automate Show.[6]

The event hosted 863 participating firms, with 610 providing detailed disclosures of their products and services. These firms represent the full spectrum of the robotics ecosystem, including robot manufacturers, sensor and component suppliers, software developers, system integrators, and service providers specialising in testing and maintenance. This diversity emphasizes the complexity of the robotics value chain, where innovation depends on collaboration across specialised domains. For example, companies like Universal Robots showcased collaborative robots (cobots) equipped with AI-driven vision systems for flexible assembly lines, while firms like SICK AG presented LiDAR sensors for precise environmental mapping, illustrating the interplay of hardware and software in modern robotics.

A prominent trend at the Automate Show is the widespread adoption of AI to enhance robotic capabilities. As shown in Figure 1, 308 out of the 610 exhibitors (50.5%) featured AI-enhanced products or services, including machine learning algorithms for motion planning and computer vision for object recognition. For example, ABB exhibited robots designed for autonomous material handling in warehouses, leveraging AI to navigate dynamic environments, while KUKA showcased robotic arms with AI-driven quality inspection systems for manufacturing precision. This presence of AI-driven solutions signals a shift away from traditional, hardware-centric (and mostly fixed) robotics to adaptable systems capable of learning patterns from data and responding to complex settings, from manufacturing plants and distribution centers to construction and mining sites.

[5] See i.e., https://ifr.org/worldrobotics/.

[6] We collect information on exhibitors from the official website of Automate Show 2024, held in Chicago IL, US.

The competitive landscape in robotics is also evolving. Alongside established global robotics leaders from Japan and Germany, such as FANUC, Yaskawa, or Siemens — who continue to advance precision automation and manufacturing systems —, the industry is witnessing the rise of software-centric entrants; these offer AI solutions that integrate with hardware machinery to expand robots' capabilities and flexibility. For instance, firms such as Intrinsic, GreyOrange, and Plus One Robotics specialise in AI-driven solutions for robotic automation, logistics, and vision-guided picking systems. This 'softwarisation' of robotics signals a new phase in industry competition, where expertise in AI, data processing, and control software increasingly drives differentiation, value creation, and value capture.

Furthermore, these shifts reflect a broader trend toward platform-based and ecosystem-oriented innovation. The robotics sector is no longer defined solely by mechanical engineering excellence but by the ability to orchestrate hardware, software, data, and integration services into cohesive solutions. This interdependence opens new niches for specialised actors, from AI algorithm developers to system integrators and Internet-of-Things providers. Both startups and global incumbents are adapting to this new landscape, competing and collaborating across different layers of the robotics value chain.

Overall, information on exhibitors from Automate Show 2024 reveals a robotics industry on the brink of structural transformation, and characterised by a complex, collaborative ecosystem and a growing emphasis on AI-driven, adaptable systems. These trends underscore the potential for robotics to meet evolving industrial and commercial demands, setting the stage for further exploration of embodied AI and its implications for the industry.

2.2 Embodied AI and the arrival of VLAMs

As AI becomes the quintessential 'large technical system' enabling innovation in complementary technologies and domains [Vannuccini and Prytkova, 2024], an important direction of AI innovation is *embodied AI* [Duan et al., 2022]. Unlike traditional AI software, embodied AI integrates algorithms with the physical environment (real-world or simulated) to apply learning models to tasks such as manipulation and navigation. Alongside increasing portability of AI at the edge [Gill et al., 2025], the field of embodied AI has witnessed significant advancements, reflecting a growing demand for AI systems that can engage dynamically with their surroundings and human agents.

As mentioned, robotics is a very fertile ground for embodied AI applications. All three key areas of robotics — perception, planning, and actuation — can potentially migrate under AI-engines. In fact, a key trend in embodied AI is the integration of natural language processing (NLP) and robotic systems, bridging semantics and physical execution [Tellex

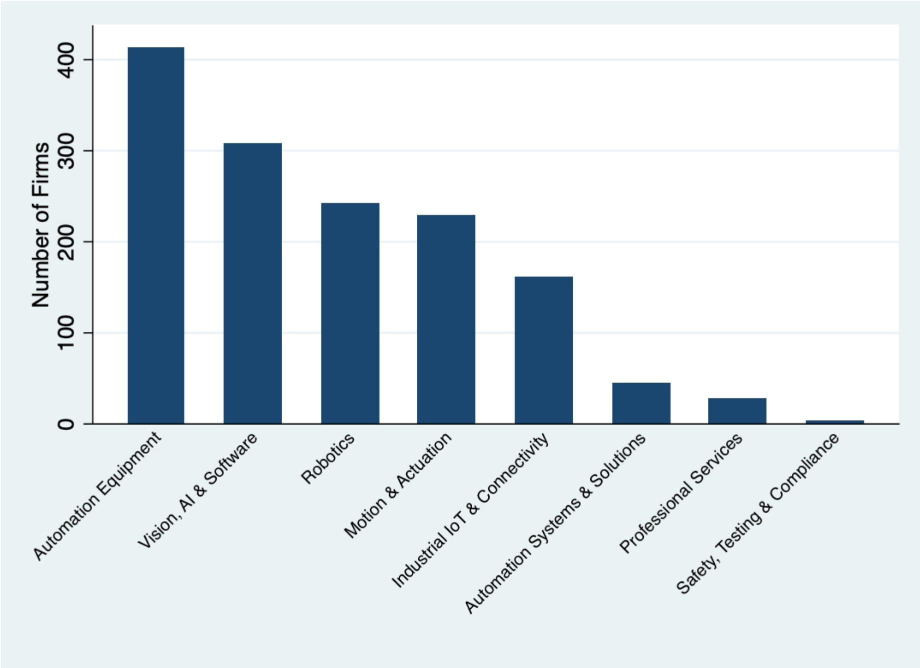


Figure 1: **Robotics Industry Overview: Automate Show 2024** This figure illustrates the distribution of industry sectors within robotics that participated in the Automate Show 2024.

et al., 2020]. Robots have interacted with language through 'grounded language' — the interpretation of natural language that situates meaning in required actions in the physical environment[7]. Research traditions on NLP and robotics are well-established; however, the arrival of and advances in LLMs have accelerated innovation in the two fields.

Our object of analysis, VLAMs, are a radical innovation at the intersection of NLP and robotics. The general idea behind VLAMs is 'upgrading' the NLP-engine of robots with LLMs. The diffusion of Transformer-based architectures, which has conquered the status of dominant design in the industry, has played a crucial role in this evolution [Brohan et al., 2023]. The set of capabilities that LLMs have acquired in recent years, multimodality in particular, allow to ease or circumvent a series of bottlenecks that have slowed down advances in robotics. For instance, grounding robotic affordances in LLMs-processed natural language opens the opportunity to train robots on the vast set of text corpora available on the Internet. The promise of VLAMs is that generalist LLMs, adequately fine-tuned for

[7]One among many examples, one, taken from Tellex et al. [2020], is the correspondence between the natural language instruction "turn left" and the grounded language interpretation "Contra-rotate the steering actuators"

embodied applications, can help achieve generalist (i.e., flexible) robots.

Rapid advancements in embodied AI carried with them the signal of potential commercial returns; in turn, this has attracted attention and substantial investments, further raising expectations for the field. Furthermore, as profitability in software AI continues to be negligible, while at the same time LLMs seems having reached a stable architectural stage, the decision to allocate resources to embodied AI in robotics appears to be a reasonable option to explore alternative channels of value capture.

Large tech companies have started exploring the intersection of AI and robotics, however following different approaches. For instance, NVIDIA's opted for developing simulation environments: their Project GR00T accelerates the development of AI-driven robots and autonomous systems by generating synthetic, physics-based data, allowing developers to train and refine models in scalable virtual environments [8].

VLAMs, which we unpack next, are an example of innovation in a related direction. We focus on this technology and on the steps taken by a key actor in the field — Google DeepMind — by virtue of the specific trajectory of evolution of this particular family of innovations, that we can rationalise with established theoretical frameworks, as well as of their technological coherence, which allows us to bound the technology neatly for analysis.

VLAMs are "are a class of multimodal models within the field of embodied AI, designed to process information from vision, language, and action modalities." [Ma et al., 2024, p.1]. VLAMs can be considered an innovation in the domain of NLP-driven robotics, as well as an broadening of vision-language models (VLMs, see Zhang et al. [2024]), usually focused on vision, that now add robotic actuation executed after processing natural language [Black et al.]. VLAMs can be seen as a step in the direction of integrating all robotics functions under the overarching principle of semantics-driven processing; in other words, to introduce end-to-end solutions for robotics grounded in a common logic. As an innovation pushing for convergence and building on novel approaches, we can consider VLAMs an emerging technology for robotics Rotolo et al. [2015]. As an emerging (sub-)field, multiple terms have been used to describe the kind of advancements we are interesting in. While Robotic foundation models might be one of the most appropriate descriptor, some studies refer to these models as Embodied Foundation Models, highlighting their role as pretrained, general-purpose AI systems that can be fine-tuned for robotic applications [Ahn et al., 2024]. The term Embodied Multimodal Agents has also gained traction in discussions surrounding AI systems that integrate multiple sensory and control modalities for real-world interaction [Yang et al.,

[8] https://nvidianews.nvidia.com/news/nvidia-powers-humanoid-robot-industry-with-cloud-to-robot-computing-platforms-for-physical-ai

2024]. In the robotics and reinforcement learning communities, terms such as World Models or Policy Learning Models have been used to describe similar efforts focused on training AI systems in simulated environments before deploying them in physical settings [Georgiev et al., 2025]. Additionally, some researchers categorize these advancements under Interactive AI or Situated AI, emphasizing the real-time, context-aware nature of these systems [Krishna et al., 2022]. The growing variety of terminology reflects the rapid expansion of research in this area, with different communities emphasizing specific technical or conceptual aspects of embodied AI integration.

A key driver of VLAMs development has been the involvement of leading technology firms actively shaping the trajectory of embodied AI. These have been exploiting economies of scope resulting from their sizeable investments on the trajectory of scaling language models with growing compute, data, and neural networks' size (parameters). Among these actors, Google DeepMind has played a central role by integrating LLMs, VLMs, and code-based policies to enhance robotic capabilities. The development by DeepMind's VLAMs research follows a structured approach, addressing fundamental challenges in planning, perception, and actuation. In Figure 2, we report a timeline of DeepMind innovation in VLAMs to illustrate the progressive integration of the field, with LLMs 'conquering' all robotics functions in the most recent iteration of the technology (RT-2).[9] All the models we discuss are rather recent, and appeared in a relative short time-period. The burst of clustered innovation in VLAMs seconds the idea that the technology is yet emerging and currently experiences a phase of ferment.

The VLAM-breakthrough began in 2022 with *SayCan*, which replaced traditional robotic planners with LLMs, enabling robots to interpret verbal instructions and plan tasks semantically [Ahn et al., 2022]. This was followed by *Inner Monologue*, which integrated VLMs to enhance visual perception and real-time responsiveness, allowing robots to perceive and react to their surroundings dynamically [Huang et al., 2022]. DeepMind further advanced VLAMs with *Code as Policies*, shifting from conventional robotic controllers to code-generating language models, enabling robots to execute high-level instructions as flexible, adaptive code [Liang et al., 2023]. In 2023, the introduction of *PaLM-E* marked a significant milestone by fusing perception and planning, combining vision encoders with LLMs to process multimodal inputs seamlessly [Driess et al., 2023]. This advancement allowed robots to analyse their environment and strategise actions simultaneously, bridging the gap

[9]This figure is based on a keynote given by Vincent Vanhoucke, who was leading Google DeepMind robotics team, at NVIDIA GTC 2024. For more details, see: https://www.nvidia.com/en-us/on-demand/session/gtc24-s61182/.

between cognitive reasoning and situational awareness. With *RT-1*, DeepMind streamlined robotic learning by replacing modular architectures with an end-to-end transformer model, improving adaptability through large-scale dataset training [Brohan et al., 2022]. The latest iteration, *RT-2*, builds upon this foundation by enhancing generalization across varied and unstructured environments, demonstrating the transition of VLAMs from experimental AI to real-world robotics [Brohan et al., 2023]. Even more recently (not included in the timeline), Google refined further its end-to-end VLAMs approach by moving its LLM-based robotics engine to their workhorse AI model, Gemini.[10] The success of Google DeepMind's systematic development of VLAMs has impressed an acceleration to the field in a rather short time, with an array of new, competing VLAMs appearing regularly [Wen et al., 2025].

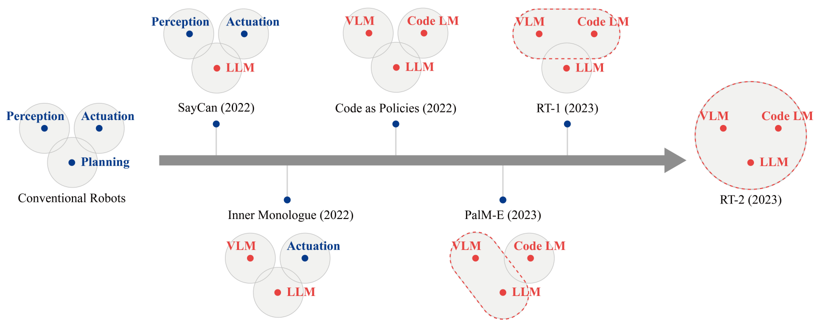


Figure 2: **Timeline of VLAM Development by Google DeepMind** This figure illustrates the major milestones in the evolution of VLAMs developed by Google DeepMind. Red-colored terms indicate key technological changes, while red dashed lines represent the integration of multiple technologies during development.

It must be stressed that VLAMs are only one of the trajectories in the recent convergence between AI and robotics. As already mentioned, due to the emerging nature of the technology, a complex landscape is forming, with competing generalist and specialised solutions. For instance, breakthrough advances are occurring also in the space of simulation models — so-called environments or 'playgrounds' — to train policies directly on the software side [Zakka et al., 2025].

[10] https://deepmind.google/discover/blog/gemini-robotics-brings-ai-into-the-physical-world/.

3 Research Questions, Methodology, and Analysis

We consider VLAMs a breakthrough and a potential rejuvenation shock to innovation in the robotics industry. For this reason, we are interested in mapping different aspects of VLAMs' emergence. Given the novelty of the technology and the infancy of its market segment, we cannot exploit longitudinal information; instead, we can compare structural *snapshots* of different areas involved in VLAM developments. Our goal is to derive general insights through a systematic mapping exercise. These insights may inform the construction of alternative scenarios for the future of AI and robotics industries, and can add to the repertoire of evidence of strategic behaviors by gatekeeper firms in technologically dynamic environments.

We opted for a three-pronged approach, that connects scientometric and market mappings, and strategy analysis. The overarching research question informing the analysis can be framed as: *Does the introduction of VLAMs mark a new phase in robotics innovation and industry dynamics?*. For all areas of analysis, we outline further, more specific research questions we decided to tackle. Each subsection that follows is structured in a consistent manner: we introduce the sub-research question(s); we illustrate the data we collected; and we present and discuss evidence from our analysis.

3.1 Scientometric Analysis

In the scientometric part of the analysis, we address the question: *What are the characteristics of the VLAMs technological trajectory?*. We are interested in understanding who develops the technology and in the modalities with which VLAMs-related knowledge is being accumulated. We tackle the question by identifying the actors involved in VLAMs publication activities in terms of country and organization type (e.g., small/large, incumbent/entrant, public/private), and by tracking the changing variety of research themes spurred by key research contributions.

3.1.1 Data

To analyse the development of VLAMs research, we employ a scientometric approach, that allows us to systematically track disciplinary distribution across different fields and identify the actors involved in VLAMs publication activities. Since capturing research specific to VLAMs technologies is challenging due to the convergence of different fields and multiple topical overlaps, we opted for a snowballing strategy for publication collection. First, we focus on a curated seed sample of 102 foundational VLAMs papers identified by Ma et al.

[2024]. We rely on this expert knowledge to bound the VLAM field. Then, we expand this sample using the OpenAlex database, tracing the seed papers' forward citations to identify additional follow-on VLAMs-related publications. This process yields 8,526 forward citations. From this data, we classify the involved actors by country and organisational type (private or public). To ensure data accuracy, we perform organizational name harmonization, using an LLM-based name cleaning process, followed by manual verification.[11] This refinement enables a precise examination of publication trends and institutional interactions within the VLAMs domain.

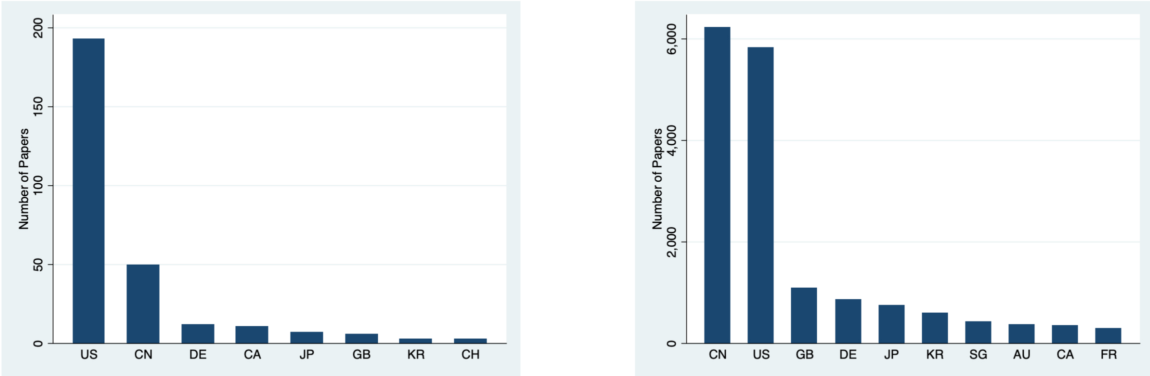
3.1.2 Evidence

In this section, we analyse geographical leadership, corporate dominance, institutional engagement, and commercialization potential in the VLAMs domain.

First, the early development of VLAMs technology was spearheaded by the United States (US), which led the field in terms of both the volume and foundational significance of research publications. As shown in Figure 3a, which captures the distribution of seed papers, the US played a central role in shaping the initial trajectory of the field. These early contributions were closely linked to broader advances in AI and robotics, underpinned by the US' robust ecosystem of research universities and technology firms. However, the landscape is rapidly evolving. Figure 3b, which presents data on forward citations, indicates that China has rapidly and substantially increased its research activity in the domain. The rising volume of citations to VLAMs-related studies produced by China suggests a concerted effort to build capacity and influence in this emerging field. While the US retains a strong lead in foundational research, China's accelerating engagement points to an increasingly competitive global research environment around VLAMs.

Second, among private actors, Google, including its subsidiary DeepMind, stands out as the most influential contributor to VLAMs research, producing the highest number of foundational publications in the field. Other major US tech giants, including Microsoft, NVIDIA, Meta, and Amazon, also maintain a visible presence, although their output remains lower relative to Google, as shown in Figure 4a. While US-based firms continue to dominate overall research activity, Chinese firms such as Huawei, Tencent, and Alibaba have markedly expanded their involvement in the field. As illustrated in Figure 4b, these companies are rapidly increasing their publication rates, signaling a phase of accelerated catch-up and growing competitiveness. This trend suggests that the corporate landscape of

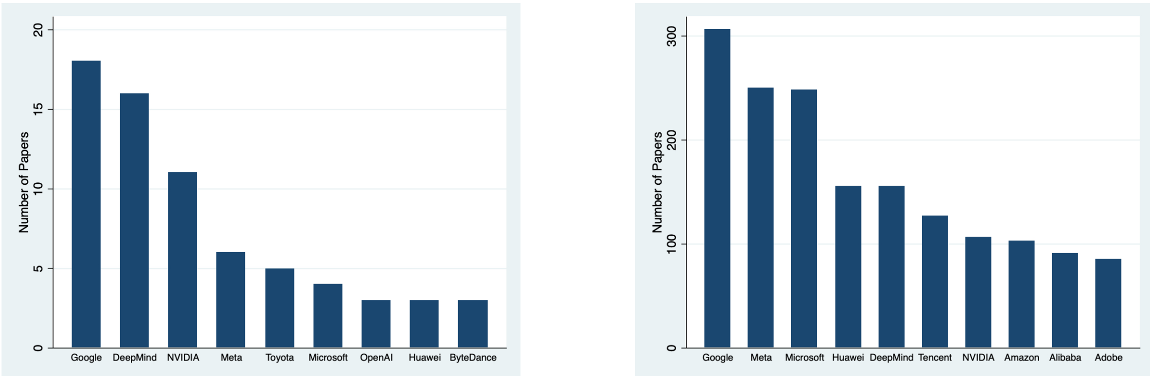
[11] We use OpenAI's GPT4o, via API access.



(a) Seed Papers (b) Forward Citations

Figure 3: **Global Trends in VLAMs Research Publications by Country** This figure illustrates (a) the distribution of seed papers and (b) the forward citations of those seed papers, categorized by country.

VLAMs research is becoming increasingly contested, with Chinese firms actively seeking to narrow the gap with their US counterparts and assert a stronger role in shaping the future trajectory of VLAMs.

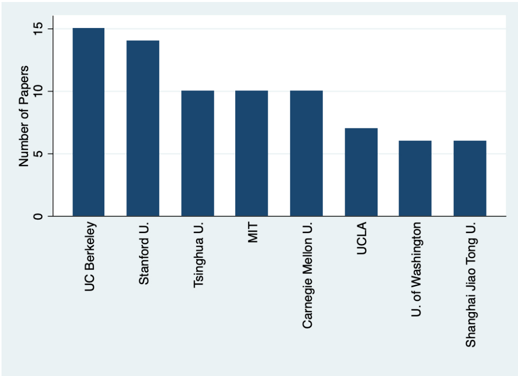


(a) Seed Papers (b) Forward Citations

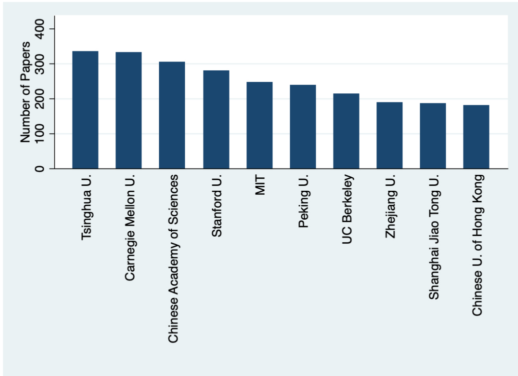
Figure 4: **VLAMs Research Publications by Private Corporations** This figure illustrates (a) the distribution of seed papers and (b) the forward citations of those seed papers, categorized by private corporation.

Third, US public institutions collectively remain the most prolific contributors to VLAMs research, particularly in terms of foundational publications. Universities such as UC Berkeley, Stanford, MIT, and Carnegie Mellon continue to play a central role in shaping the early development of the field (Figure 5a). At the same time, Chinese public institutions, notably

Tsinghua University, have significantly scaled up their research activities. This growing engagement is further reflected in the broader body of follow-up work, where institutions such as the Chinese Academy of Sciences, Peking University, and Zhejiang University have become increasingly prominent (Figure 5b). While US institutions maintain their lead in absolute terms, the rising subsequent development by Chinese universities highlights a broader trend toward global diffusion and intensified international competition in the development of VLAMs technologies.



(a) Seed Papers

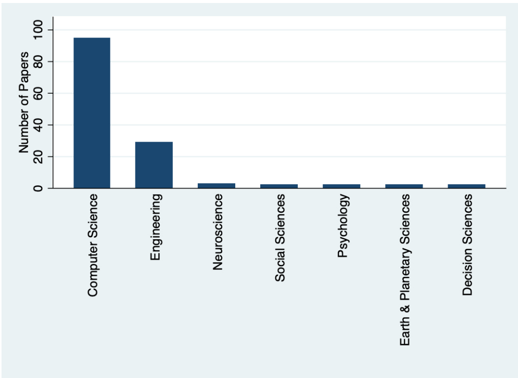


(b) Forward Citations

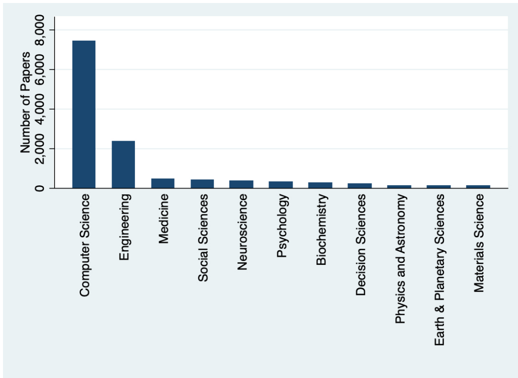
Figure 5: **VLAMs Research Publications by Public Institutions** This figure illustrates (a) the distribution of seed papers and (b) the forward citations of those seed papers, categorized by Public Institution.

Fourth, a closer examination of the subdomains within VLAMs research reveals Google's strong and, in most cases, dominant position across the key areas that define VLAMs systems: constituents of VLAMs, control policy, and task planning, as outlined in Ma et al. [2024] classification scheme. In foundational work (Appendix B1a, B1b, and B1c), UC Berkeley leads in the constituents subdomain, with Google ranking second. However, in both control policy and task planning, Google dominates the early research landscape. This leadership becomes even more pronounced in the broader body of follow-up research, where Google maintains a leading position across all three subdomains, as shown in Appendix B2a, B2b, and B2c. While other firms and institutions contribute meaningfully to the development of VLAMs systems, Google's consistent presence across both foundational and follow-on research highlights its central role in shaping the direction of the field and reinforces its strong research footprint in the evolving VLAMs ecosystem.

Fifth, the disciplinary distribution of VLAMs research indicates that foundational work remains heavily concentrated in computer science (Figure 6a). However, the subsequent development landscape reveals a notable shift toward interdisciplinary engagement, with VLAMs-related research increasingly used in engineering, medicine, social sciences, and biochemistry (Figure 6b). This broadening scope underscores the growing relevance of VLAMs beyond theoretical and algorithmic work, as researchers across fields begin to explore their application in real-world contexts. The diffusion of VLAMs research into such diverse domains provides early but significant signals of commercial potential, suggesting that these technologies are beginning to move from the lab into practice. This transition marks a critical step in the evolution of VLAMs from a nascent research frontier to a technology of cross-sector relevance and emerging industrial interest.



(a) Seed Papers



(b) Forward Citations

Figure 6: **Disciplinary Distribution of VLAMs Research Publications** This figure illustrates (a) the distribution of seed papers and (b) the forward citations of those seed papers, categorized by research field.

To complement the descriptive analysis and better understand the evolution and structure of research on VLAMs, we conducted a network analysis of the key actors and research domains shaping the field. The seed paper network shows a bimodal degree distribution, with a large number of nodes having few connections and a small core of highly connected actors. This pattern reflects a centralised structure, where early research activities were concentrated among a few key players, such as Google, Microsoft, UC Berkeley, serving as hubs and gatekeepers within the knowledge network. In contrast, the forward citation network exhibits a heavily skewed distribution with a long tail, suggesting a broader, though less interconnected, field expansion. While a few institutions, including Google, Meta, Tencent, and Carnegie Mellon University, maintain high connectivity, the majority of new contributors remain at the periphery, reflecting the growing dispersion and collaborative ties of VLAMs research (Figure 7).

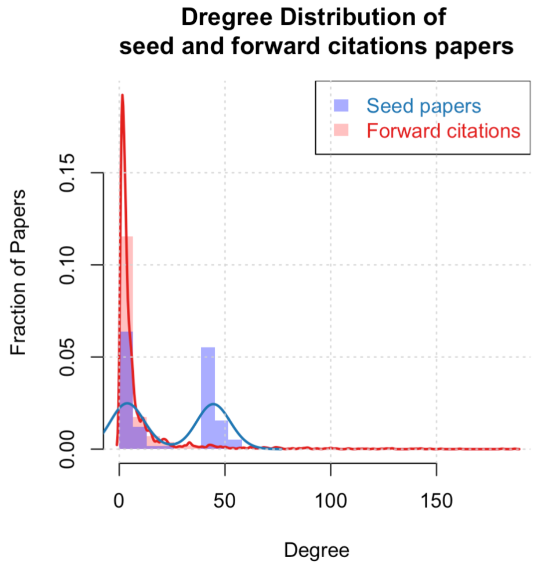
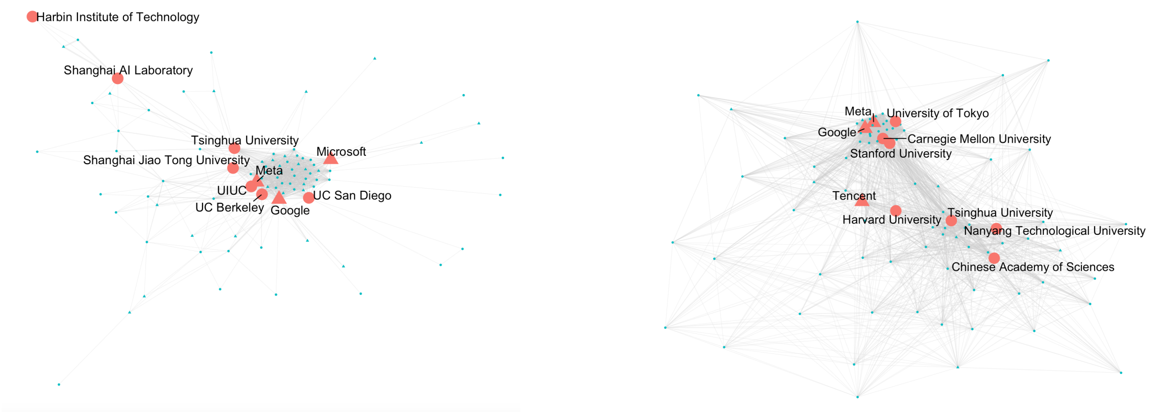


Figure 7: **Degree Distribution in VLAMs Research** This figure shows the structural evolution of the field.

The top contributors in terms of co-authorship centrality further illustrate this evolution. The early phase is marked by a denser, more centralised network, indicating a tightly connected core of entities, including universities such as UC Berkeley, Tsinghua University, and Google, that initiated the field with high levels of collaboration and, given their high degree of interconnectedness, acted as gatekeepers in the early development of VLAMs (Figure 8a). We calculated standard centrality indicators and found a relatively high network density of 0.27 and an average betweenness centrality of 0.14, indicating strong internal connectivity and the presence of key intermediary nodes facilitating collaboration flows. In contrast, the forward-citation network presents a markedly lower density of 0.0035 and betweenness centrality of 0.043, indicating a broader, more fragmented expansion phase. This shift suggests a diversification of the field: a larger number of actors are citing and building upon early work, but are less interconnected (Figure 8b). This reflects the diffusion of VLAMs concepts into a wider range of institutions. However, as found in the descriptive analysis, the main actors remain concentrated in the US and China and — in the private sector — continue to be large tech companies diversifying into robotics.

The thematic structure of the VLAMs network reinforces the evolutionary dynamics observed. The seed paper network is tightly clustered around core AI, computer vision, and control engineering topics, indicating a foundational focus on technical integration (Fig-

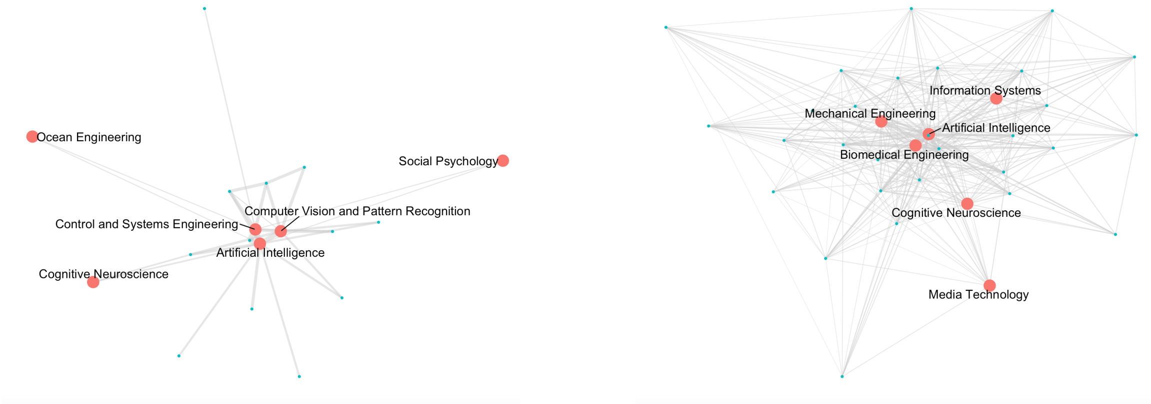


(a) Seed Papers (b) Forward Citations

Figure 8: **Co-authorship Networks in VLAMs Research** This figure illustrates co-authorship network in (a) seed papers and (b) the forward citations of those seed papers. Nodes represent institutions, with circles indicating public institutions and triangles indicating private companies.

ure 9a). This is supported by a relatively high network density $(0,1428)$ and an average betweenness centrality of $0,3458$, reflecting a cohesive and centralised knowledge structure during the early development phase. In contrast, the forward citation network displays a more fragmented thematic landscape, with research expanding into areas such as biomedical engineering, information systems, and molecular biology (Figure 9b). This shift is accompanied by a decrease in network density $(0,0769)$ and betweenness centrality $(0,2778)$, indicating a more distributed structure with more points of control of the knowledge flows. This thematic diversification highlights the growing interdisciplinary engagement with VLAMs technologies and signals a transition from technical feasibility to practical (and potentially commercial) application across sectors.

Finally, a key trend in the evolution of VLAMs research is the shifting dynamic between corporate and academic contributions (Figure 10a). The initial breakthroughs in the field were largely driven by private firms such as Google, often in collaboration with public-sector institutions. Over time, public institutions began catching-up in VLAM fundamental research (Figure 10b). This trajectory, which we discuss more at length in Section 4 may echo historical precedents such as the Bell Labs, where broader academic engagement trailed breakthrough research initiated by private labs [Gertner, 2013]. It goes without saying that we cannot exclude the possibility that the prominence of private actors in developing early VLAM research descends from pre-existing fundamental public research done in related



(a) Seed Papers (b) Forward Citations

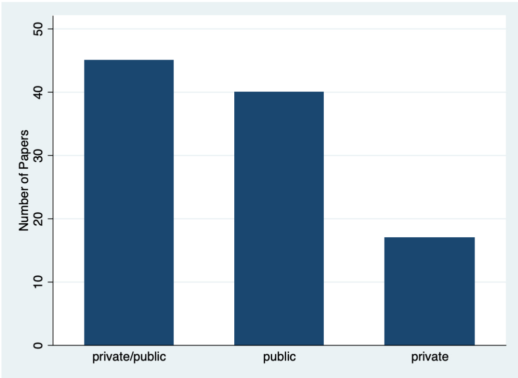
Figure 9: **Thematic Networks of VLAMs Research** This figure illustrates thematic structure of (a) seed papers and (b) the forward citations of those seed papers. Nodes represent institutions, with circles indicating public institutions and triangles indicating private companies.

technologies (e.g., VLMs) — research that our sample does not capture. However, as we consider VLAMs a radical innovation in the field, our insight retain its importance: a breakthrough in robotics might emerge from a private labs, signaling a (at least localised) important shift in the regime of knowledge generation.

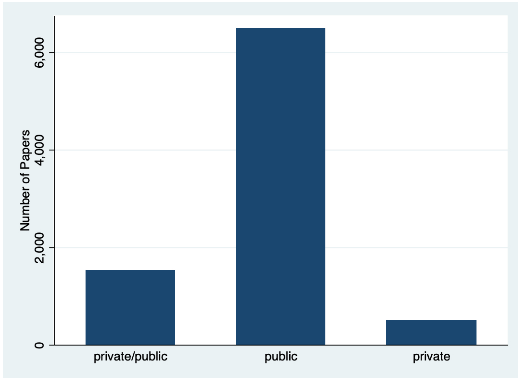
Our idea find support in a more general analysis about the different types of strategies pursued by Big Tech: in fact, Google's approach to academic engagement aligns with what Rikap [2024] refer to as a 'University Strategy' — a structured effort to shape knowledge production by conducting fundamental research in-house while at the same time funding university research and retaining control over core technological assets (Appendix B3a and B3b).

3.2 Market Analysis

In this part, we address the question *who is riding the new wave of AI-powered robotics?*. We are interested in understanding whether this emerging market segment is experiencing softwarisation; that is, whether the major players in the convergence of AI and robotics are established robotics companies that pursue the exploration of a new niche, or entrants from the software industry. Furthermore, we look at whether there is overlap between market actors and those appearing in our scientometric mapping.



(a) Seed Papers



(b) Forward Citations

Figure 10: **Collaboration Patterns in VLAMs Research Publications** This figure illustrates (a) the distribution of seed papers and (b) the forward citations of those seed papers, categorized by collaboration type: private-public collaboration, private-only, or public-only.

3.2.1 Data

To understand the features of the emerging market segment around VLAMs, we gathered and pooled data from a variety of complementary sources. First, we manually compiled a list of established firms and startups specialising in LLM-powered robotics. The list was compiled by collecting information on business descriptions from news outlets, LinkedIn profiles, and robotics-related trade events, using the major terms used to identify products and developments in the field (e.g., robotic or world foundation models) as keywords. Then, we leveraged the database Crunchbase to gather detailed information about these organizations' business activities, including funding rounds, strategic alliances, and mergers and acquisitions (M&A).

3.2.2 Evidence

We explore the characteristics of the actors driving the innovation in VLAMs (or VLAM-related) technology. Table 1 (in Appendix A) contains all the information we compiled on our market sample.[12] A first set of columns present the strategies of the listed companies, highlighting their collaborations with dominant firms and their funding approaches. Approximately half of the companies on the list are startups that have secured funding from various sources, including venture capital, while others have been backed or led by domi-

[12]Last update of Table 1: March 2025.

nant tech firms. In terms of collaborations, many of the listed companies are relying on the software development and simulation labs of Google DeepMind and NVIDIA to accelerate their progress in embodied AI. A notable trend is the emergence of startups such as SkildAI, TRobotics, and Physical Intelligence that focus exclusively on software development, specifically on foundation models for hardware companies in the robotics field. Another notable trend involves collaborations with major retailers and warehouse companies, as these sectors increasingly turn to robotics to improve automation and logistics.

The second part of the table shows the type of technology these companies are developing, specifically VLAMs models, and whether they are working in robotics or are focusing on the development of humanoid robots, the most recent trend in the industry. Among the companies designing hardware solutions, more than half are developing humanoid robots with VLAMs technology, and some are integrating software solutions from major companies. For example, NVIDIA launched the already mentioned Project GR00T, a general-purpose foundation model designed exclusively for humanoid robots.

We also examine the overlap between actors actively involved in scientific and market developments, and find that the intersection remains limited. First, when comparing firms active in VLAMs market developments with those participating in the Automate Show 2024 (Section 2.1), only a few companies (e.g., Intrinsic and Neural Robotics) are present in both lists. This suggests that many of the firms advancing VLAMs technologies may not yet be embedded in the traditional industrial robotics ecosystem typically represented at major trade shows, indicating a gap between legacy automation players and new entrants focused on embodied AI. Second, the overlap between scientific contributors (Section 3.1) and market development actors in the VLAMs market is similarly narrow. While a small number of software-focused firms involved in foundational research also appear in scientific literature, the majority of emerging market players operate outside formal academic channels. This likely reflects the early stage of market formation, in which commercial efforts are still emerging and firms are focused on exploring applications rather than academic dissemination. As a remark, we cannot exclude that the limited overlaps is a byproduct of our construction of the actors lists; however, our evidence is suggestive of a degree of disconnect between established robotics players and new actors introducing VLAM-related solutions — a trace of potential upcoming market disruption.

Together, these findings suggest that the VLAMs domain is being shaped by a series of parallel dynamics: on the one hand, we witness the entry of a distinct new cohort of actors — neither fully rooted in the established robotics industry nor strongly linked to the academic research base. The limited crossover between scientific and market development actors points to the emergence of a new wave of embodied AI firms operating at the periphery of existing institutional structures. These actors are usually vertically specialised (e.g., in software development). On the other hand, we see efforts geared towards vertical integration, both by large tech platforms and by entrants partnering with firms active in manufacturing and distribution channels. As the field evolves, greater convergence may unfold; in line with stylised developments of industries and market segments, we might expect waves of consolidation through acquisition as soon as disruption and turbulence settles down, as the technology becomes fully technologically reliable and economically viable.

3.3 Strategy Analysis

In this third mapping exercise, we address the question: *Are the actions taken by the technology pioneer indicative of an attempt to gatekeep the industry?*. We explore the specific case of Google DeepMind, by virtue of its key role of innovator in VLAMs, as evidenced by our scientometric analysis. We examine whether its recent actions indicate an intention to consolidate influence and establish control over the innovation trajectory in the VLAMs domain.[13]

3.3.1 Data

To explore Google DeepMind's strategic action in the VLAMs domain, we conduct a qualitative analysis of their key publications as primary data sources. Specifically, we use their more recent publications, such as those related to the Open X-Embodiment initiative [14], the OpenVLA project [15], and Gemini Robotics [16] as qualitative datasets.

3.3.2 Evidence

A key objective in the path towards achieving industry dominance involves establishing a dominant design, which significantly raises entry barriers for potential competitors. Dominant designs enable firms to shape both the direction and production of technology, effectively gatekeeping and steering future innovation. For robotics, the key bottleneck in developing a dominant design lies in the creation of generalist robotic policies applicable

[13] Google DeepMind's protagonism in the field is proved by the fact that the very label VLAM has been introduced for the first time in the companies' research papers — what we can consider as an attempt at naming, and thus shaping, the field.

[14] https://robotics-transformer-x.github.io/

[15] https://openvla.github.io/

[16] https://deepmind.google/discover/blog/gemini-robotics-brings-ai-into-the-physical-world/

across numerous embodiments — a task heavily reliant on diverse and extensive datasets. In the next paragraphs, we trace Google DeepMind moves to achieve dominance and set the standard for VLAMs. This case study offers useful insights on the steps a company that aims at shaping and controlling a specific technology field can take. We identify four different moves.

First, Google DeepMind strategically addressed the key bottleneck in the development of the technology by championing the creation of an extensive pooled datasets that aggregates diverse robotic learning scenarios. The Open X-Embodiment initiative[17] exemplifies this strategic move: it supported the creation of extensive robotic learning datasets (called RT-X) that position numerous external research laboratories as complementors rather than competitors. By addressing the heavy data diversity requirements for robotics research, DeepMind strategically transforms the broader research ecosystem into collaborators, enhancing its own central role and influence. As we will see in Section 4, this move is typical of platformisation strategies.

A second move has been that of pursuing the dissemination of Google's foundational VLAMs through open-sourcing. Strategic opening [Alexy et al., 2018] is put in practice with the OpenVLA project[18], which builds upon open-source models such as Meta's Llama2. By providing an accessible, open-source backbone model with OpenVLA, DeepMind promotes wide adoption and standardisation of its technological framework across the robotics industry. This strategy further solidifies its potential position as an influential gatekeeper, as it controls critical components of the innovation pipeline.

Third, after having signalled a willingness to support open-source, the company moved to consolidate its strategic position further, migrating and integrating its VLAMs programme under the umbrella of Gemini Robotics framework. In its original form, Gemini Robotics extends Gemini 2.0 model's capabilities by incorporating physical actions as an output modality, enabling direct robot control for diverse real-world tasks. Strategic partnerships with the company Apptronik to develop next-generation humanoid robots, and collaborations with partners and testers such as Agile Robots, Agility Robotics, Boston Dynamics, and Enchanted Tools, amplify DeepMind's ecosystem influence. These proprietary advancements and expanding partnerships position DeepMind at the forefront of generalist robotic development, fostering dependency among external actors and strengthening its control over the innovation pipeline, thus solidifying its role as a technological gatekeeper.

In summary, Google DeepMind's strategic approach can be characterised by a series of

[17] See https://robotics-transformer-x.github.io/.

[18] See https://openvla.github.io/.

progressive moves: consolidating data to remove the fundamental reverse salient in the field; establishing a broad collaboration network that is directed to function as complementor to core technological developments; pushing for the adoption of open-source models; re-integration into an end-to-end proprietary solution. These moves can be rationalised from a strategic innovation and management perspective — we dedicate to that part of the next Section.

4 Conceptual Framework and Discussion

Our analysis offered a cross-domain analysis of VLAMs that identified key actors in science and commercial development. We also qualitatively identified the strategic levers employed by a pioneer of the technology. In this Section, we generalise our findings to derive insights that can be useful beyond the VLAMs case. First, we situate VLAMs within the context of breakthrough and radical innovation. Second, we elaborate on the role of private vs public labs in developing the technology. Third, we discuss whether the strategic decision pursued in the field reproduce expected outcomes.

4.1 VLAMs as Breakthrough Innovation

We consider VLAMs as a case of breakthrough innovation. Breakthrough, as it displays the features of an emerging technology [Rotolo et al., 2015] while also producing ripple effects on knowledge production and follow-on technological innovation, as well as on market dynamics. We discussed the potential market implications of VLAMs as enablers of lateral entry, disruption and shift in the beneficiaries of value capture in the industry in subsection 3.2. Here, we focus on the broader insights that the VLAMs case offers for innovation theory. In general, the technology seems to have a large scope for improvement, given that it appears to be following scaling laws (that is, an exponentional relationship between resources allocated to its development and performance), with a learning curve possibly steeper than that of non-embodied language models [Sartor and Thompson, 2024].

The notable fact of VLAMs development is that innovation occurs through integration or 'unification' of separate aspects of robotics under the common umbrella of semantics, through LLMs. Advances and execution of perception, planning and actuation converge under the same unifying principle, and replace their established software 'engine' with LLMs by virtue of these models' increasing multimodal capabilities. As we mentioned, this process is akin to what W. Brian Arthur called redomaining [Arthur, 2009]: the functions of robotics remain the same, but they are ported under a different working logic. One can witness a similar dynamics in AI, where specific business modules inside a firm replace 'classic' software backbones with neural-network-based ones, which execute the same function but operate according to different principles (that of data-driven prediction) [Bresnahan, 2019]. The useful feature of redomaning is that, by reframing bottlenecks, it can offer a different angle at problem-solving activities and unlock opportunities that were previously hidden. This technological 'translation' echoes historical instances where innovations from one domain transform another, such as digital signal processing in telecommunications [Licklider, 1960].

More generally, the establishment of VLAMs captures a peculiar way in technological change unfolds. A key model to capture the evolution of technology is the one inspired by Darwinian principles: innovation as a process of variation, selection, and retention [Grodal et al., 2023]. In our case, variation occurs in an upstream domain (software AI); is then implemented in a narrow downstream application (hardware robotics); the progress in achieving generalist robots is retained and induce a widening of the scope of VLAM research (as we identified in the scientometric analysis) as well as the development of more integrated models. More than a case of evolution, it appears as a case of *co-evolution*, as theorised by Almudi et al. [2024], between software and hardware domains. Furthermore, the dynamics we outline is well-approximated by models of innovation percolation across complex technological spaces [Silverberg and Verspagen, 2005].

Finally, besides being radical in a technological sense — redomaining changes the very underlying principle driving advances in robotics — VLAMs can be disruptive innovations as well. Disruption might not be limited to a reversal of fortune in the actors that capture value in the market segment (and the industry more generally), but rather extend to the regime of knowledge production that lead to innovation in the field, where public labs research may disrupt established fundamental science conducted at public institutions. We discuss this next.

4.2 The Role of Corporate-led Innovation: Bell Labs 2.0?

The development of VLAMs highlights the role of corporate actors in steering foundational research in embodied AI. While Google DeepMind's VLAMs build on an articulate, pre-existing VLM backbone, our scientometric and strategy analyses have shown that large tech companies purposefully decided to take the helm of innovation in the field. From the earliest integration of LLMs into robotic planning systems (as in SayCan), to the recent deployment of transformer models such as RT-2 or OpenVLA, the trajectory of VLAMs has been profoundly shaped by a handful of firms.

This pattern recalls historical precedents in mid-twentieth-century technological development: in particular, the role of Bell Labs [Gertner, 2013]. Bell Labs was widely regarded as a model of corporate research, credited for foundational innovations such as the transistor, laser, and early computing architectures. Over the years, Bell Labs employed researchers that contributed or initiated key fields in basic science (Claude Shannon, who developed information theory, is a notable example). The labs operated as a vertically integrated research and development hub within AT&T, maintaining a long-term perspective on scientific inquiry while directly contributing to the firm's technological innovations. This mode of knowledge generation, which we label 'Bell Lab 1.0', captures how corporations — in specific historical phases — have been able to leverage their vast resources to undertake and accelerate innovations.

The comparison with Bell Labs, however, is not merely historical. It reflects a possible reconfiguration of the innovation system, wherein corporate entities are reclaiming a role in foundational research. As shown in our findings, key VLAMs papers have been authored by private sector actors, while follow-on research is increasingly concentrated within public institutions. In effect, the expected direction of knowledge flow — where academic research seeds commercial applications — is somehow reversed in the VLAM case. This dynamics aligns with observations that Big Tech, with their access to proprietary infrastructure (e.g., computational resources) and 'deep financial pockets', are exerting growing pressure on public institutions, outpacing them in high-impact AI research, while simultaneously favouring a 'brain drain' across academic rangs, from PhD students to professors [Ahmed et al., 2023]. Today's Google Deepmind's activism in the VLAMs space maps the Bell Labs' vertically-integrated, basic science-driven approach to high-tech innovation. For this reason, we advance the idea that the regime of innovation described here represents a 'Bell Labs 2.0' model.

The importance of corporate research (at least in the US) seems to have grown historically when universities were unable to contribute frontier research [Arora et al., 2024]. This is possibly the case for embodied AI, at least in part. However, the renewed protagonism of corporate research does not stem necessarily from a reaction to sluggish research in universities. Instead, it might reflect the desire of controlling a promising market segment (as we highlighted in the strategy analysis in subsection 3.3), as well as an attempt at recouping sizeable sunk investments in software AI by pushing advances in AI application sectors that will hopefully yield economic returns.

Regardless of its real driver, the pattern we detected suggests that a new division of knowledge labour could emerge, one in which firms retain frontier knowledge while universities operate as complementary sites of refinement, diffusion, and adaptation. Understanding

VLAMs as an instance of Bell Labs 2.0 allows us to illustrate the evolving role of corporations in the face of emergence of breakthrough technologies that enable the convergence and integration of different domains. In particular, the Bell Labs 2.0 idea emphasizes a shift in the innovation landscape, where corporate actors not only drive technological breakthroughs but also shape the direction of public research.

4.3 Platformisation, Open-source, and the Orchestration of Innovation

The strategic behaviour of leading firms in the VLAMs domain can be understood through the lens of platformisation and open-sourcing strategies. In emerging technology fields, dominant actors often pursue strategies that allow them to shape technological trajectories while consolidating control over critical infrastructure and standards. These strategies are not merely about technological leadership, but about constructing and orchestrating ecosystems where innovation becomes increasingly dependent on proprietary resources, protocols, or architectures maintained by the platform leader.

In the case of VLAMs, the approach adopted by Google DeepMind exemplifies this mode of strategic orchestration. DeepMind's platform-based approach to VLAMs includes the provision of open-access models and datasets, encouraging external development while positioning itself as a central hub. The Open X-Embodiment initiative, discussed in subsection 3.3 for instance, aggregates diverse robotic learning datasets, positioning external research labs as complementors rather than competitors. Similarly, the OpenVLA project, built on open-source frameworks, promotes standardisation of VLAMs architectures, fostering widespread adoption. This orchestration strategy mirrors platformisation dynamics, in which control over data flows and ecosystem standards yields competitive advantages. By enabling third-party developers to build on its platforms, Google DeepMind enhances network effects, as more participants contribute to and depend on its technological framework, reinforcing its gatekeeper role.

Platformisation can serve multiple strategic purposes. First, it enables the firm to centralise control over data flows, which are essential for the continued improvement of large-scale embodied models. As external developers and institutions adopt DeepMind's VLAMs platforms, they contribute data, refinements, and usage insights that flow back into the firm's core models. Second, platformisation fosters network effects: as more actors adopt the platform's standards and interfaces, switching costs for alternative solutions increase, thereby reinforcing the incumbent's dominance. Third, by encouraging innovation around a common infrastructure, the platform leader effectively positions itself as the gatekeeper of the technological ecosystem, capable of influencing both the pace and direction of subsequent advancements.

In this context, open-sourcing is an important strategy to gain dominance in a market. Alexy et al. [2018] discuss the notion of 'strategic opening', and Gray Widder et al. [2024] highlight the motives driving for-profit tech companies to use this strategy. Besides opening to increase user-base with the aim of achieving control of a platform (e.g., the case of Google's Android), open-source investments can be done to harm competitors that benefit from proprietary technology. For instance, IBM invested over a billion dollars in the development of Linux operating system to challenge Microsoft and its Windows OS. Alternatively, firms might open-source a technology while purposefully shift the source of value capture (a strategy known as value migration): this is the case of cloud computing, where firms like Amazon and Microsoft provide open-access infrastructure (e.g., AWS, Azure) to attract developers, but maintain dominance through proprietary tools and data analytics. Similarly, a firm can open-source a solution to monetise it as a service. In the VLAMs domain, DeepMind's platformisation seems to go in line with at least some of these motives: pre-empting commercial competitors from developing and lock users in proprietary solutions; securing market control through dataset aggregation and model standardisation. These parallels suggest that platformisation in VLAMs could lead to concentrated market power, but also accelerate innovation by lowering entry barriers for smaller firms if they position themselves as complementors.

Strategic opening can be a permanent or a temporary solution. For instance, the transition of Google DeepMind's robotic foundation models under the Gemini product family might be an early warning signal of a return to a proprietary model regime, after a position of dominance in the technological platform has been achieved. At the same time, the original decision to open-source can spur the entry of alternatives that can challenge the gatekeeper. This is precisely what has happened in the VLAMs space, with the emergence of a series of open models for robotics collected in the model markeplace HuggingFace.[19]

Eventually, understanding these strategies within the theoretical frameworks of platformisation and ecosystem governance provides an insightful lens for interpreting current developments in the VLAMs domain and potential long-term competitive outcomes. It can also help managers and policy makers to form an awareness of the profound ongoing transformation of robotics, as the forces shaping organisational and innovation practice are poised to evolve; established hardware firms and software entrants will face different sets of opportunities and constraints in robotics, to be supported or eased with different types of actions.

[19]See https://huggingface.co/lerobot.

5 Conclusion and Future Research

5.1 Conclusion

We have examined the emergence of VLAMs as a breakthrough innovation in embodied AI and robotics by tracking ongoing dynamics in three distinct but complementary domains: scientific, market, and strategy. To the authors' knowledge, this is the first study offering a systematic mapping of the impacts produced by the structural integration of AI models with the technology underlying an established sector — in this case, robotics. We considered VLAMs a radical as well as a disruptive innovation by virtue of (i) its 'redomaining' of all fundamental robots' functions — perception, planning, and actuation — under a semantic-based unifying principle; and (ii) its shifting in the locus of value creation in the market towards VLAM-producing (software) firms.

More in details, the scientometric analysis reveals that initial research in this area has been driven largely by corporate actors with public institutions increasingly contributing to subsequent advancements. This pattern suggests that what we labelled a 'Bell Labs 2.0' dynamic is at work in this segment of robotics, where foundational knowledge is produced in private research labs before expanding into academic exploration.

The market analysis showed that the commercialisation of VLAMs is characterised by a mix of incumbents and new entrants, with dominant technology firms leveraging their expertise in AI (LLMs) to consolidate their positions. Many robotics startups are aligning with these major players through partnerships and licensing agreements, suggesting that the VLAMs market is heading towards more concentration and vertical integration. Overall, software firms seem to be at the roots of value generation in the VLAMs and embodied-AI robotics, indicating a trends toward the softwarisation of the industry that parallels developments in other sectors, such as automotive.

Finally, strategic analysis highlighted that firms leading the development of VLAMs are employing platformisation strategies, using open-source releases to influence industry standards while simultaneously reinforcing their own competitive advantages. This approach has enabled them to position themselves as technological gatekeepers, setting the trajectory for VLAMs adoption across multiple sectors.

We used our findings to generalise these domain-specific insights to overall considerations that can apply to industries exposed to rejuvenation by rapid technological change enabled by advances in AI.

5.2 Future Research

Our technology of interest, VLAMs, is extremely novel — we identified it as an emerging technology. Its novelty constrained our main data source to publications as early signals. However, publications may not fully capture the underlying technologies in this field. Future research could complement scientometric analyses with additional data sources, such as patent filings, to provide a more systematic investigation of technological advancements and technological positioning among key actors. VLAMs-related patents are virtually non-existent at this stage, given the structural time lag characterising patent application and granting. However, future work could leverage this additional source of information.

Furthermore, the scope of this study is inherently bounded to the early phase of VLAMs development. While this provides valuable insight into the initial dynamics of science, market, and strategy, longitudinal studies are needed to trace how the field evolves over time. In particular, further research could examine shifts in contributions and patterns such as strategic alliance or M&A among key players. In addition, examining the evolving roles of leading firms across different VLAMs domains—such as control policy or task planning would clarify their influence on market dynamics and competitive structures. Such an approach would offer a more refined understanding of the evolving technological fields, and of the competitive and cooperative dynamics shaping the development of the VLAMs market.

Finally, follow-on studies may investigate the sectoral dynamics of VLAMs adoption, with particular attention to how firms in traditional industries respond to the growing availability of embodied AI capabilities. As VLAMs-powered systems transition from laboratory experiments to real-world applications, sectors such as automotive, logistics, and manufacturing are poised to become important sites of technological experimentation and strategic deployment. Exploring how these industries experiment with VLAMs technologies, including the challenges of implementation could provide insights into the practical implications of VLAMs and their role in driving industry-specific innovation.

References

W Brian Arthur. *The nature of technology: What it is and how it evolves.* Simon and Schuster, 2009.

Sebastian Sartor and Neil Thompson. Neural scaling laws in robotics. *arXiv preprint arXiv:2405.14005*, 2024.

Zongwei Liu, Wang Zhang, and Fuquan Zhao. Impact, challenges and prospect of software-defined vehicles. *Automotive Innovation*, 5(2):180–194, 2022.

Simone Vannuccini and Ekaterina Prytkova. Artificial intelligence's new clothes? a system technology perspective. *Journal of Information Technology*, 39(2):317–338, 2024.

Jiafei Duan, Samson Yu, Hui Li Tan, Hongyuan Zhu, and Cheston Tan. A survey of embodied ai: From simulators to research tasks. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 6(2):230–244, 2022. doi: 10.1109/TETCI.2022.3141105.

Sukhpal Singh Gill, Muhammed Golec, Jianmin Hu, Minxian Xu, Junhui Du, Huaming Wu, Guneet Kaur Walia, Subramaniam Subramanian Murugesan, Babar Ali, Mohit Kumar, et al. Edge ai: A taxonomy, systematic review and future directions. *Cluster Computing*, 28(1):18, 2025.

Stefanie Tellex, Nakul Gopalan, Hadas Kress-Gazit, and Cynthia Matuszek. Robots that use language. *Annual Review of Control, Robotics, and Autonomous Systems*, 3(1):25–55, 2020.

Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Xi Chen, Krzysztof Choromanski, Tianli Ding, Danny Driess, Avinava Dubey, Chelsea Finn, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. *arXiv preprint arXiv:2307.15818*, 2023.

Yueen Ma, Zixing Song, Yuzheng Zhuang, Jianye Hao, and Irwin King. A survey on vision-language-action models for embodied ai, 2024. URL `https://arxiv.org/abs/2405.14093`.

Jingyi Zhang, Jiaxing Huang, Sheng Jin, and Shijian Lu. Vision-language models for vision tasks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46 (8):5625–5644, 2024. doi: 10.1109/TPAMI.2024.3369699.

Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, et al. $\pi0$: A vision-language-action flow model for general robot control. corr, abs/2410.24164, 2024. doi: 10.48550. *arXiv preprint ARXIV.2410.24164*.

Daniele Rotolo, Diana Hicks, and Ben R Martin. What is an emerging technology? *Research policy*, 44(10):1827–1843, 2015.

Michael Ahn, Debidatta Dwibedi, Chelsea Finn, Montse Gonzalez Arenas, Keerthana Gopalakrishnan, Karol Hausman, Brian Ichter, Alex Irpan, Nikhil Joshi, Ryan Julian, et al. Autort: Embodied foundation models for large scale orchestration of robotic agents. *arXiv preprint arXiv:2401.12963*, 2024.

Yijun Yang, Tianyi Zhou, Kanxue Li, Dapeng Tao, Lusong Li, Li Shen, Xiaodong He, Jing Jiang, and Yuhui Shi. Embodied multi-modal agent trained by an llm from a parallel textworld. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 26275–26285, 2024.

Ignat Georgiev, Varun Giridhar, Nicklas Hansen, and Animesh Garg. PWM: Policy learning with multi-task world models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL `https://openreview.net/forum?id=hOELrZfg0J`.

Ranjay Krishna, Donsuk Lee, Li Fei-Fei, and Michael S Bernstein. Socially situated artificial intelligence enables learning from human interaction. *Proceedings of the National Academy of Sciences*, 119(39):e2115730119, 2022.

Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, et al. Do as i can, not as i say: Grounding language in robotic affordances. *arXiv preprint arXiv:2204.01691*, 2022.

Wenlong Huang, Fei Xia, Ted Xiao, Harris Chan, Jacky Liang, Pete Florence, Andy Zeng, Jonathan Tompson, Igor Mordatch, Yevgen Chebotar, et al. Inner monologue: Embodied reasoning through planning with language models. *arXiv preprint arXiv:2207.05608*, 2022.

Jacky Liang, Wenlong Huang, Fei Xia, Peng Xu, Karol Hausman, Brian Ichter, Pete Florence, and Andy Zeng. Code as policies: Language model programs for embodied control. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9493–9500. IEEE, 2023.

Danny Driess, Fei Xia, Mehdi S. M. Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, Wenlong Huang, Yevgen Chebotar, Pierre Sermanet, Daniel Duckworth, Sergey Levine, Vincent Vanhoucke, Karol Hausman, Marc Toussaint, Klaus Greff, Andy Zeng, Igor Mordatch, and Pete Florence. Palm-e: An embodied multimodal language model. In *arXiv preprint arXiv:2303.03378*, 2023.

Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Joseph Dabis, Chelsea Finn, Keerthana Gopalakrishnan, Karol Hausman, Alex Herzog, Jasmine Hsu, et al. Rt-1: Robotics transformer for real-world control at scale. *arXiv preprint arXiv:2212.06817*, 2022.

Junjie Wen, Yichen Zhu, Jinming Li, Minjie Zhu, Zhibin Tang, Kun Wu, Zhiyuan Xu, Ning Liu, Ran Cheng, Chaomin Shen, et al. Tinyvla: Towards fast, data-efficient vision-language-action models for robotic manipulation. *IEEE Robotics and Automation Letters*, 2025.

Kevin Zakka, Baruch Tabanpour, Qiayuan Liao, Mustafa Haiderbhai, Samuel Holt, Jing Yuan Luo, Arthur Allshire, Erik Frey, Koushil Sreenath, Lueder A. Kahrs, Carmelo Sferrazza, Yuval Tassa, and Pieter Abbeel. Mujoco playground, 2025. URL `https://arxiv.org/abs/2502.08844`.

Jon Gertner. *The idea factory: Bell Labs and the great age of American innovation*. Penguin, 2013.

Cecilia Rikap. Varieties of corporate innovation systems and their interplay with global and national systems: Amazon, facebook, google and microsoft's strategies to produce and appropriate artificial intelligence. *Review of International Political Economy*, 31(6): 1735–1763, 2024.

Oliver Alexy, Joel West, Helge Klapper, and Markus Reitzig. Surrendering control to gain advantage: Reconciling openness and the resource-based view of the firm. *Strategic Management Journal*, 39(6):1704–1727, 2018.

Timothy Bresnahan. Artificial intelligence technologies and aggregate growth prospects. *Prospects for economic growth in the United States*, pages 132–172, 2019.

Joseph CR Licklider. Man-computer symbiosis. *IRE transactions on human factors in electronics*, (1):4–11, 1960.

Stine Grodal, Anders Dahl Krabbe, and Mia Chang-Zunino. The evolution of technology. *Academy of management annals*, 17(1):141–180, 2023.

Isabel Almudi, Francisco Fatas-Villafranca, John Foster, and Jason Potts. Coevolution and dynamic processes: an introduction to this issue and avenues for future research. *Review of Evolutionary Political Economy*, 5(3):399–423, 2024.

Gerald Silverberg and Bart Verspagen. A percolation model of innovation in complex technology spaces. *Journal of Economic Dynamics and Control*, 29(1-2):225–244, 2005.

Nur Ahmed, Muntasir Wahed, and Neil C Thompson. The growing influence of industry in ai research. *Science*, 379(6635):884–886, 2023.

Ashish Arora, Sharon Belenzon, Konstantin Kosenko, Jungkyu Suh, and Yishay Yafeh. The rise of scientific research in corporate america. *Organization Science*, 2024.

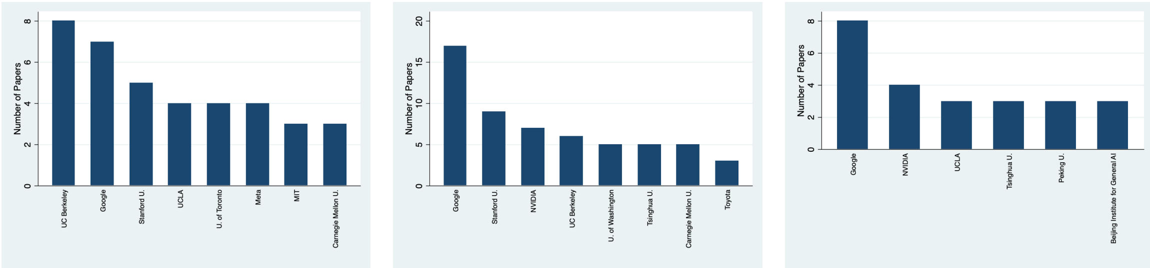
David Gray Widder, Meredith Whittaker, and Sarah Myers West. Why 'open' ai systems are actually closed, and why this matters. *Nature*, 635:827–833, November 2024. doi: 10.1038/s41586-024-08141-1. URL `https://doi.org/10.1038/s41586-024-08141-1`. Published online 27 November 2024.

A VLAMs startups sample

Company	Year	Strategy			Development		Technology		Funding	Paper Contributions	
		Type of Funding	Parent	Collaborator	Softw.	Hardw.	AI Model	Humanoid		Fwd Cit.	Seed
Google DeepMind	2010		Google		X	X	X	X		710	16
Covariant	2017	Licensed by Amazon		Otto Group	X		X		222	6	0
NVIDIA	1993				X		X			707	11
META	2021				X		X			850	6
Figure	2022			OpenAI, Microsoft, Nvidia, BMW	X	X	X	X	854	0	0
Sanctuary AI	2018	Different investors		Microsoft	X	X	X	X	30	0	0
Collaborative Robotics	2022	Different investors		NVIDIA		X	X		140	0	0
Agility Robotics	2015	Different investors		Amazon, Ricoh, Tompkins, Zion, Manhattan	X	X	X	X	178	0	0
Apptronik	2004			Google DeepMind, GXO, Mercedes Benz, NVIDIA		X	X	X	28	0	0
Agibot	2023	Different investors		Lingang Group	X	X	X	X	844	0	0
Intrinsic	2021	Owned by Alphabet			X		X			532	2
Mentee Robotics	2022	Different investors		NVIDIA	X	X	X	X	17	0	0
Sereact	2021	Different investors		Daimler Truck, Ludwig Meister	X	X	X			0	0
Fourier	2015	Funding led by SoftBank		China Construction Bank	X	X	X	X		0	0
1X	2014	Funding led by OpenAI		OpenAI	X	X	X	X		0	0
XPENG Robotics	2016	Different investors	XPENG		X	X	X			0	0
Agile Robots	2018	Different investors			X	X	X			0	0
GALBOT	2023	Different investors		Peking University	X	X	X	X		0	0
Hillbot	2024	Different investors			X		X			0	0
Intbot AI	2025			NVIDIA		X	X	X		0	0
Neura Robotics	2019			Han's Robot	X	X	X	X		0	0
Skild AI	2023	Funding led by Amazon			X		X			0	0
LimX	2005	Different investors		NVIDIA	X	X	X	X		0	0
DeepRoute.ai	2019			NVIDIA	X	X	X			0	0
TRobotics	2024	Different investors		ABB	X		X			0	0
Physical Intelligence	2024	Different investors led by OpenAI, Amazon			X		X			3	1
Tesla	2003				X	X	X	X		0	0
Oversonic	2020				X	X	X	X		0	0

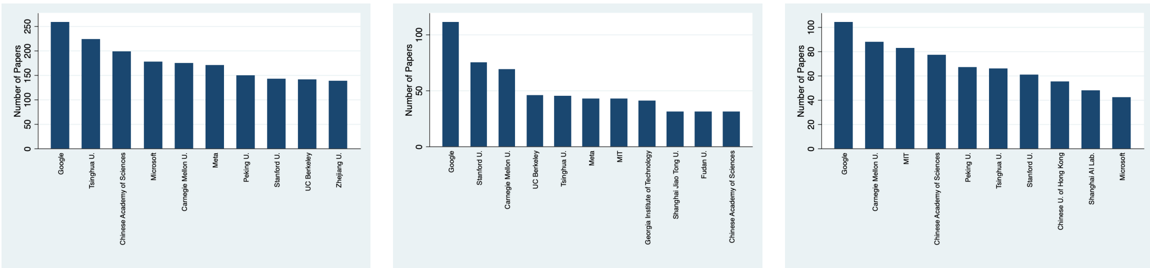
Table 1: Overview of startup companies developing VLAMs

B Additional Scientometric descriptives



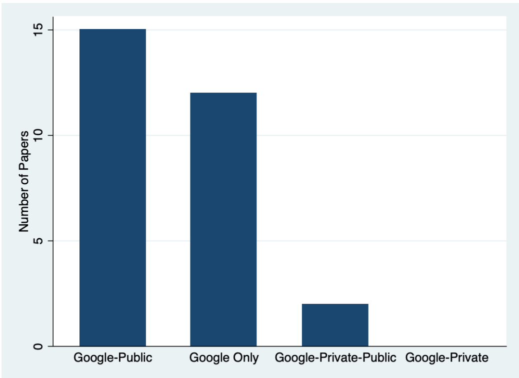
(a) Constituents of VLAMs (b) Control Policy (c) Task Planner

Figure B1: **Subdomains of VLAMs in Seed Paper Publications** This figure illustrates the distribution of seed papers across three subdomains of VLAMs research: (a) Constituents of VLAMs, (b) Control Policy, and (c) Task Planner.

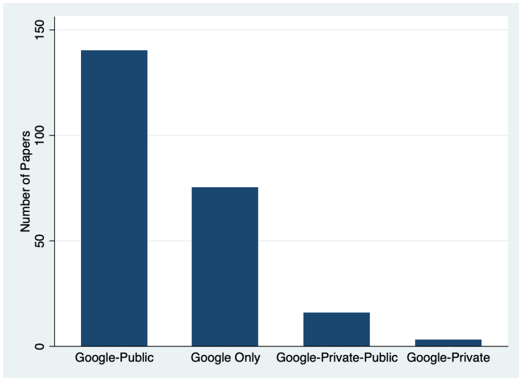


(a) Constituents of VLAMs (b) Control Policy (c) Task Planner

Figure B2: **Subdomains of VLAMs in Forward Citation Publications** This figure illustrates the distribution of forward citations of seed papers across three VLAMs subdomains: (d) Constituents of VLAMs, (e) Control Policy, and (f) Task Planner.



(a) Seed Papers



(b) Forward Citations

Figure B3: **Collaboration Patterns of Google in VLAMs Research Publications** This figure illustrates (a) the distribution of seed papers and (b) the forward citations of those seed papers, categorized by Google's collaboration type: Google only, Google–public, Google–private, and Google–private–public collaborations.

DOCUMENTS DE TRAVAIL GREDEG PARUS EN 2025
GREDEG Working Papers Released in 2025

2025-01 Bruno Deffains & Frédéric Marty
Generative Artificial Intelligence and Revolution of Market for Legal Services

2025-02 Annie L. Cot & Muriel Dal Pont Legrand
"Making war to war" or How to Train Elites about European Economic Ideas: Keynes's Articles Published in L'Europe Nouvelle during the Interwar Period

2025-03 Thierry Kirat & Frédéric Marty
Political Capitalism and Constitutional Doctrine. Originalism in the U.S. Federal Courts

2025-04 Laurent Bailly, Rania Belgaied, Thomas Jobert & Benjamin Montmartin
The Socioeconomic Determinants of Pandemics: A Spatial Methodological Approach with Evidence from COVID-19 in Nice, France

2025-05 Samuel De La Cruz Solal
Co-design of Behavioural Public Policies: Epistemic Promises and Challenges

2025-06 Jérôme Ballet, Damien Bazin, Frédéric Thomas & François-Régis Mahieu
Social Justice: The Missing Link in Sustainable Development

2025-07 Jérôme Ballet & Damien Bazin
Revoir notre manière d'habiter le monde. Pour un croisement de trois mouvements de pensée: capabilités, services écosystémiques, communs

2025-08 Frédéric Marty
Application des DMA et DSA en France : analyse de l'architecture institutionnelle et des dynamiques de régulation

2025-09 Éléonore Dodivers & Ismaël Rafaï
Uncovering the Fairness of AI: Exploring Focal Point, Inequality Aversion, and Altruism in ChatGPT's Dictator Game Decisions

2025-10 Hayk Sargsyan, Aleksandr Grigoryan & Olivier Bruno
Deposits Market Exclusion and the Emergence of Premium Banks

2025-11 Lubica Stiblarova & Anna Tykhonenko
Talent vs. Hard Work: On the Heterogeneous Role of Human Capital in FDI Across EU Member States

2025-12 Mathieu Chevrier
Social Reputation as one of the Key Driver of AI Over-Reliance: An Experimental Test with ChatGPT-3.5

2025-13 Rania Belgaied
L'accessibilité aux médecins généralistes libéraux pour la ville de Nice

2025-14 Lorenzo Corno, Giovanni Dosi & Luigi Marengo
Behaviours and Learning in Complex Evolving Economies

2025-15 Grégory Donnat
Real Exchange Rate and External Public Debt in Emerging and Developing Countries

2025-16 Michela Chessa
Politics as A (Very) Complex System: A New Methodological Approach to Studying Fragmentation within a Council

2025-17 Mathieu Lambotte & Benjamin Montmartin
Competition, Conformism and the Voluntary Adoption of Policies Designed to Freeze Prices

2025-18 Leonardo Ciambezi & Alessandro Pietropaoli
Relative Price Shocks and Inequality: Evidence from Italy

2025-19 Matteo Orlandini, Sebastiano Michele Zema, Mauro Napoletano & Giorgio Fagiolo
A Network Approach to Volatility Diffusion and Forecasting in Global Financial Markets

2025-20 Daniele Colombo & Francesco Toni
Understanding Gas Price Shocks: Elasticities, Volatility and Macroeconomic Transmission

2025-21 Simone Vannuccini
Move Fast and Integrate Things: The Making of a European Industrial Policy for Artificial Intelligence

2025-22 Patrice Bougette, Oliver Budzinski & Frédéric Marty
Revisiting Behavioral Merger Remedies in Turbulent Markets: A Framework for Dynamic Competition

2025-23 Amal Ben Khaled, Rami Haj Kacem & Nathalie Lazaric
Assessing the Role of Governance and Environmental Taxes in Driving Energy Transitions: Evidence from High-income Countries

2025-24 Sandye Gloria
Emergence. Another Look at the Mengerian Theory of Money

2025-25 Barbara Brixová, Anna Tykhonenko, Ľubica Štiblárová & Marianna Siničáková
Middle Income Convergence Trap Phenomenon in CEE Countries

2025-26 Grégory Donnat, Maxime Menuet, Alexandru Minea & Patrick Villieu
Does Public Debt Impair Total Factor Productivity?

2025-27 Nathalie Lazaric, Loubna Echajari & Dorota Leszczyńska
The Multiplicity of Paths to Sustainability, Grand Challenges and Routine Changes: The Long Road for Bordeaux Winemakers

2025-28 Olivier Brette & Nathalie Lazaric
Some Milestones for an Evolutionary-Institutional Approach to the Circular Economy Transition

2025-29 Pierre Boutros, Eliana Diodati, Michele Pezzoni & Fabiana Visentin
Does Training in AI Affect PhD Students' Careers? Evidence from France

2025-30 Luca Bargna, Davide La Torre, Rosario Maggistro & Benjamin Montmartin
Balancing Health and Sustainability: Optimizing Investments in Organic vs. Conventional Agriculture Through Pesticide Reduction

2025-31 Jean-Luc Gaffard
Equilibre général et évolution: un éclairage théorique

2025-32 Patrice Bougette, Frédéric Marty & Simone Vannuccini
Competition Law Enforcement in Dynamic Markets: Proposing a Flexible Trade-off between Fines and Behavioural Injunctions

2025-33 Taekyun Kim, Gabriela Fuentes & Simone Vannuccini
When Robots Met Language (Models): An Exploration of Science and Strategy in the vision-language-action Models Space