OPEN FORUM

Challenges as catalysts: how Waymo's Open Dataset Challenges shape AI development

Sam Hind[1] · Fernando N. van der Vlist[2] · Max Kanderske[3]

Abstract

Artificial intelligence (AI) and machine learning (ML) are becoming increasingly significant areas of research for scholars in science and technology studies (STS) and media studies. In March 2020, Waymo, Google/Alphabet's autonomous vehicle project, introduced the 'Open Dataset Virtual Challenge', an annual competition leveraging their Waymo Open Dataset. This freely accessible dataset comprises annotated autonomous vehicle data from their own Waymo vehicles. Yearly, Waymo has continued to host iterations of this challenge, inviting teams of computer scientists to tackle evolving machine learning and vision problems using Google's data and tools. This article analyses these challenges, situating them within the context of the 'Grand Challenges' of artificial intelligence (AI), which aimed to foster accountable and commercially viable advancements in the late 1980s. Through two exploratory workshops, we adopted a 'technographic' approach to examine the pivotal role of challenges in the development and political economy of AI. Serving as an organising principle for the AI innovation ecosystem, the challenge connects companies and external collaborators, driving advancements in specific machine vision domains. By exploring six key themes—interface methods, incrementalism, metrics, AI vernacular, applied domains, and competitive advantages—the article illustrates the role of these challenges in shaping AI research and development. By unpacking the dynamic interaction between data, computation, and labour, these challenges serve as catalysts propelling advancements towards self-driving technologies. The study reveals how challenges have historically and presently shaped the evolving landscape of self-driving and AI technologies.

Keywords Autonomous driving · Machine vision · Challenges · Waymo · AI ecosystem · Incrementalism

1 Introduction

Artificial intelligence (AI) and machine learning (ML) are becoming increasingly significant areas of research for scholars in science and technology studies (STS) and media studies. Scholars are exploring various aspects, including labour considerations (Tubaro et al. 2020) and the politics of algorithmic decision-making (Sánchez-Monedero and Dencik 2022) to the materiality of computation (Rella 2023), the role of training datasets (Thylstrup 2022), and the economic underpinnings of AI ethics (Steinhoff 2023). In this article, we contribute to this growing literature by studying the organisation of ML, machine vision 'challenges' used to foster technological innovation. This is particularly the case for emerging ML-dependent AI systems, such as autonomous vehicles.

Our objective is to examine how challenges shape applied AI research and development (R&D), using the case of Waymo, Google/Alphabet's autonomous vehicle project.[1] Waymo's recurring annual Open Dataset Challenges (2020–23) represent one example of open competitions organised for the global ML and data science community.[2] Our investigation into these challenges adopts a material approach bridging digital STS (Vertesi and Ribes

✉ Sam Hind
sam.hind@manchester.ac.uk

Fernando N. van der Vlist
f.n.vandervlist@uva.nl

Max Kanderske
max.kanderske@uni-siegen.de

1 University of Manchester, Manchester, UK

2 University of Amsterdam, Amsterdam, Netherlands

3 University of Siegen, Siegen, Germany

---

[1] 'Company', Waymo, https://waymo.com/company/.

[2] 'Waymo Open Dataset', Waymo, https://waymo.com/open/; 'Challenges', Waymo, https://waymo.com/open/challenges/.

2019), platform studies (Helmond et al. 2019), and work on the political economy of AI (Luitse and Denkena 2021; Srnicek 2022; Van der Vlist et al. 2024), furthering insight into the phenomenon of 'platform automobility' (Hind et al. 2022; cf. Forelle 2022; Hind and Gekker 2022; Steinberg 2022) and autonomous driving (Hind 2019; Iapaolo 2023; Marres 2020; Sprenger 2022; Stilgoe 2017; Tennant and Stilgoe 2021).

Building upon a workshop conducted by the authors at the University of Siegen (Siegen, Germany), focussed on the Waymo Open Dataset, we adopt a 'technographic' approach (Bucher 2018; Van der Vlist et al. 2024) to explore how challenges play a crucial role in the development and political economy of AI and autonomous vehicles. Through a 'scavenging-style' ethnography (Seaver 2017), we examine their significance in 'convening' third-party developers (Egliston and Carter 2022 p. 10), considering how platform features, technical documentation, and other materials figure in the incremental advancement of AI systems and technologies.

Waymo has been a leader in the autonomous vehicle industry ever since it started as the Google Self-Driving Car project in 2009 (Markoff 2010). It continues to compete with car manufacturers like Tesla (through its mis-sold 'Autopilot' feature), Ford (former backer of Argo AI), and China's Baidu; other Big Tech-funded projects like Zoox (a subsidiary of Amazon), dedicated autonomous vehicle passenger service (AVPS) operators like Cruise, and chip manufacturers like NVIDIA and Mobileye. Together, they shape an industry that has entered a new, mature phase, as key players have variously consolidated their self-driving vehicle operations (Mobileye), written-off related assets (Ford), or pivoted to other autonomous vehicle domains (Aurora, self-driving trucks). Cruise's travails in San Francisco have only reiterated the difficult crossroads the industry has now reached (Biddle 2023; Hawkins 2023).

ML and data science challenges and competition-hosting platforms are numerous. Google subsidiary Kaggle, described by CEO D. Sculley as the 'rainforest of machine learning' (Pan and Fields 2022), provides a platform for users to discover and publish datasets, explore and construct models, and participate in various data science challenges to enhance their skills, earn ranking points, and win prizes.[3] The Grand Challenge platform serves as another instance of an open web-based environment for challenges, focussing specifically on the end-to-end development of ML solutions in biomedical imaging.[4] Such competitions can encompass

diverse topics or algorithmic techniques since the respective datasets and evaluation criteria are typically provided by the competition hosts. Waymo also are not the only technology company to conduct competitions centred around their own datasets and rules: Netflix previously ran the 'Netflix Prize' inviting ways to improve its algorithmic film recommendation system, Cinematch (Bennett and Lanning 2007). The competition was exclusively open to external contestants, excluding individuals affiliated with Netflix, highlighting the 'boundary work' that digital platforms undertake, as they establish and manage the parameters of such competitions (Van der Vlist 2022, p. 102; cf. Helmond et al. 2019).

This article argues that challenges serve as touchpoints or *interfaces* between companies like Waymo, deeply involved in the application domain of self-driving technology, and the applied AI/ML community, including academia and machine vision subfields. This interface is a novel development in the automotive industry, encompassing open datasets, leaderboards, arXiv and GitHub pages, Computer Vision and Pattern Recognition (CVPR) workshops, metrics, and ML methods. Ultimately, challenges and the ML techniques developed in them are intended to facilitate the 'interoperation' of AI, computer vision, and related technologies in the field (Hind 2023).

In elucidating the role of specific machine vision challenges,[5] several themes can be considered provisionally important for the critical examination of AI. First is the role of challenges as a primary 'organizing principle' (Ribes et al. 2019, p. 281) within the AI development and production pipeline. This encompasses aspects such as the provision of (training and test) datasets typically associated with ML challenges, the concentration of computing power, the implications for the externalisation of human labour in AI, and the dynamics between AI platform companies and external complementors (cf. Srnicek 2022). Second, despite the grandeur often associated with challenges and the hype surrounding AI, it is essential to acknowledge the incrementalism in AI progress. This includes the ability to measure advancements through evaluative metrics like 'Average Precision' (AP). Thirdly, challenges play a crucial role, particularly through competitions, prizes, and leaderboards, in convening third-party developers and businesses to build, capture, and ultimately 'sell' self-driving technology (cf. Egliston and Carter 2022, p. 13–14). Overall, machine vision challenges emerge as pivotal components that bridge the realms of AI/ML R&D and the practical application of AI

---

[3] 'Competitions', Kaggle, https://www.kaggle.com/competitions; 'Kaggle Progression System', Kaggle, https://www.kaggle.com/progression.

[4] Grand Challenge, https://grand-challenge.org/.

[5] Waymo's perception and prediction challenges predominantly pertain to machine or computer vision capabilities—an integral facet of ML that centres on automating information extraction from visual sensory inputs or images. This encompasses a diverse range of technologies, software, hardware products, integrated systems, methodologies, actions, and expertise.

in the realm of automobility. They operate at the intersection of science and business, playing a significant role in consolidating the dominant positions of leading companies in the field—arguably a 'domain-independent' (Ribes 2019, p. 525) blueprint for the commercialisation of AI more broadly.

In the next section, we offer a contextual understanding of challenges in technology innovation and engineering. First, we situate Waymo's challenges within the historical context of the 'Grand Challenges' era of AI in the late 1980s. We then delve into the transition towards an 'incremental' AI approach, driven by the pragmatic necessities of commercial AI development. In this context, we contend that challenges act as a pivotal organising principle, and catalyst, for AI, amalgamating vital components like training data, computational power, and expert labour, fostering collaboration and propelling AI advancement. Subsequently, we introduce our 'scavenging-style' methodology for scrutinising AI and machine vision challenges. We outline our exploratory workshops, which entailed a thorough examination of the foundational platforms and tools underpinning Waymo's Open Dataset Challenges. Moving forward, we unveil six important themes that surfaced from the workshops and our ensuing investigations. These encompass challenges as *interfaces*, the importance of *incrementalism*, the role of evaluative *metrics and benchmarks*, the *vernaculars* of AI, the allure of *applied domains*, and the pursuit of *competitive advantages*. Concluding, we emphasise the exigency for further research into ML challenges, and the wider political economy of AI/ML.

2 The history of grand challenges and AI innovation

2.1 From strategic computing to autonomous robots: AI research and DARPA's grand challenge era (1983–2007)

Challenges, competitions, and prizes have long played a role in driving technological innovation. In the late 1980s, the concept of Grand Challenges emerged as a framework for realising research in science and technology. Raj Reddy's 1988 Presidential Address to the Association for the Advancement of Artificial Intelligence (AAAI) aimed to propel AI research towards tangible outcomes. Despite 'twenty-five years of sustained support' (Reddy 1988 p. 9) from organisations such as the Defense Advanced Research Projects Agency (DARPA), National Science Foundation (NSF), and NASA, Reddy contended that AI now needed to enter into 'an era of accountability' (p. 9).

To thrive in this new era, Reddy argued that AI should 'create a vision for the future' both 'exciting and challenging' (1988, p. 17). This vision, he believed, should extend

beyond the mere 'demonstration of intelligent systems' and instead involve 'bold national initiatives' capable of '[capturing] the imagination of the public'. Reddy identified these initiatives as the Grand Challenges of AI, proposing six examples, including a 'World Champion Chess Machine' and an 'Accident Avoiding Car' (p. 18).

In Reddy's mention of an 'accident avoiding' car, he asserted that 'a new generation automobile equipped with an intelligent cruise control using sonar, laser, and vision sensors could eliminate 80–90% of … fatal accidents and cost less than 10% of the total cost of the automobile' (p. 18). Whilst still a considerable distance from such a goal, Waymo today state that accident prevention, and safety more broadly, is 'at the heart of everything we do' (Waymo 2023) with numerous safety-related reports and technical papers released each year by the company (e.g. Favarò et al. 2023).

However, what made Reddy's address even more intriguing was the looming prospect of a new 'winter' for AI research, as substantial funding cuts were being implemented by the US government. The Strategic Computing Initiative (SCI), operating from 1983 to 1993, failed to deliver anticipated advancements in 'machine intelligence' despite an additional $1 billion in initial DARPA funding (Roland and Shiman 2002). As Roland and Shiman considered, the concept of '"Grand Challenges" simply replaced the former goal, machine intelligence' whilst 'the strategy and even the tactics remained the same' (2002, p. 3).

Reddy, interestingly, does not use the term 'machine intelligence' nor mentions DARPA's SCI. Nevertheless, he clearly states the need for AI to become self-sufficient—a task it struggled to accomplish. The onset of the AI winter in the late 1980s can be seen as a direct consequence of DARPA's recalibration of AI research funding, as well as an attempt to guide computer scientists and AI researchers towards commercially viable applications of AI.

One project funded by the SCI helps to connect these narratives: the Autonomous Land Vehicle (ALV) program, which sought to operationalise prior DARPA research on machine vision. As Roland and Shiman explain, the crucial question was 'whether [the ALV] could take the next step to high-level, real-time, three-dimensional IU [image understanding]' (2002, p. 220). The ALV was intended as a 'test bed' for the SCI, inviting university and commercial partners to bid for development contracts in order to provide 'tangible evidence that the money spent on the AI program was paying off' (p. 224).

In May 1984, after Carnegie Mellon University (CMU), General Electric, Honeywell, and Columbia University had all secured contracts on the project, the ALV was publicly demonstrated for the first time, successfully navigating a 1,016-m course in 1,060 s, allegedly 100 times faster than any previous autonomous vehicle (Roland and Shiman 2002, p. 228). However, as the pressure to host more public

demonstrations took precedence over developmental interests, such as designing an integrated vision and planning system for navigation, the ALV project encountered insurmountable challenges. Competing interests, lack of technological standardisation, and interoperability issues across different systems led to the project's downfall in 1986, with CMU initiating the development of their own 'Navlab' vehicle. DARPA officially terminated the ALV project in April 1988, due to claims of a 'demo-driven' culture that had overtaken the program, and DARPA's reassessment of the SCI budget (Roland and Shiman 2002, p. 246).

Fifteen years later, in 2004, DARPA hosted the first 'great robot race' (Buehler et al. 2007), the 'DARPA Grand Challenge', in the Mojave Desert, California. From an original 106 applicants, 15 teams competed to drive a 150 mile off-road course, with the hope of winning a $1 million cash prize. Completing just 5% of the route, CMU's Red Team vehicle 'Sandstorm' travelled the furthest, still failing to win the prize money, with the competition subsequently referred to as the 'debacle in the desert' (Hooper 2004). 'It was clear then', as Buehler et al.'s (2007: IX) review of the inaugural competition concluded, 'that the challenge was indeed "grand"'. Quickly followed by a similar all-terrain challenge in 2005, won by the Stanford University team and their vehicle 'Stanley' (Thrun et al. 2007), a 2007 'Urban Challenge' pushed competitors to design an autonomous vehicle to traverse a more realistic urban terrain, comprising of traffic and intersections (DARPA 2007). A team from CMU was once again victorious, with Tartan Racing's 'Boss' vehicle completing the course in just over four hours. Arguably 'groundbreaking', it was the 'first time autonomous vehicles [had] interacted with both manned and unmanned vehicle traffic in an urban environment' (DARPA 2007). This marked the delayed initiation of the Grand Challenge era of AI by DARPA, as they began to identify commercially viable models of AI in autonomous driving and similar applications.

2.2 From vision to organising principle: challenges in commercial AI technology development (since 2009)

In recent years, there has been a shift in the landscape of AI challenges, moving away from DARPA-funded Grand Challenges (Roland and Shiman 2002) towards more incremental versions hosted by start-ups, research centres, and platform firms (Hind et al. 2022). This transition reflects Raj Reddy's goal of the Grand Challenge era: to drive commercial applications of AI and reduce reliance on state funding. The launch of Google's Self-Driving Car project in 2009, led by 2005 DARPA Grand Challenge winner, Sebastian Thrun (Markoff 2010), was the obvious transition point between these two eras.

Our contention here is that these incremental challenges serve as a primary 'organizing principle … for technology development' (Ribes et al. 2019, p. 281) within domains that desire to use AI. Following Woolgar (1985), we view these challenges as occasions to study the day-to-day activities of AI researchers, and the material traces they leave behind. Today, ML has become the dominant strand of AI, relying on training data, computing power, and (expert) labour. We contend here that challenges are the mechanism through which these components are most effectively brought together, shaping the processes and trajectory of AI technology from development to its eventual deployment. We will briefly discuss these elements of training data, computing power, and (expert) labour next.

Srnicek (2022) argues that these three components are crucial for AI production, particularly in terms of the monopoly power of Big Tech companies like Google/Alphabet in shaping AI platforms and services, or what he refers to as 'AI centralization' (Srnicek 2022). He suggests that the collection of ML training data no longer offers a competitive advantage due to the prevalence of the platform business model and the 'explosive growth of open datasets' (p. 258). Waymo's Open Dataset, with its 'nearly 17 h of video, with labelling for 22 million 2D objects and 25 million 3D objects', is highlighted as an example (p. 258). Consequently, the challenge of starting without data has become less of a concern for actors in various domains.

Yet the presence of the challenge format suggests that doing something *with* the data remains significant. As we contend here, well-annotated and voluminous training data are not always readily available, with only a few initiatives within each AI domain maintaining useful and usable datasets. Building and maintaining such (open) datasets has been critical to Waymo's autonomous driving vision. These datasets play a vital role in attracting participants to the challenge format and aligning them with internal development timelines.

Furthermore, Srnicek contends that cloud computing power is 'increasingly where AI monopolies and moats are being built' (Srnicek 2022, p. 249). He argues that this is because of the 'concentrated ownership of immense computing resources (compute) and the *systems and lures* built for attracting the small supply of high-skill workers' (p. 249–250, emphasis added). Only the largest and most well-capitalised firms can afford to develop cloud computing systems capable of training models on data-rich scenarios. As Luitse and Denkena (2021, p. 3) ask, 'who can further scale up their compute capacity'? The ability to run numerous experiments quickly and efficiently is crucial in the empirical nature of AI research, involving tasks such as 'tuning hyperparameters, testing on data from outside the training dataset, debugging any problems, and so on' (Srnicek 2022, p. 251).

It is through the optimisation of hardware such as Graphics Processing Units (GPUs) and Google's own Tensor

Processing Units (TPUs) that advances in AI are being carved out (Rella 2023). Waymo and other challenge organisers believe that external competition is the most effective means of conducting large chunks of this optimisation work. By providing computational capacity, more iterations can be performed by a larger number of teams, thereby accelerating progress.

In addition, neither data nor compute holds much value without skilled labour to leverage them. Hugely sought-after by Big Tech and AI firms, computer scientists and related graduates command substantial salaries. Open-source initiatives serve as mechanisms for channelling graduate talent into the right areas, with frameworks offering 'premade tools, libraries, and interfaces…often based on the same ones used internally by companies' (Srnicek 2022, p. 252). Challenges, therefore, serve as a primary avenue for funnelling 'new AI talent' (Luchs et al. 2023, p. 9), equipping graduates with the necessary skills to work with these pre-existing tools and interfaces and engaging them in pre-defined problems chosen by the toolmakers (Steinhoff 2022; Luchs et al. 2023). If these frameworks 'become feeder networks for the emerging generations of talent' (Srnicek 2022, p. 253), challenges can be seen as talent contests, pitting the best new talent against their peers. Piecemeal 'micro-work', routinely used to prepare training data for ML work (Tubaro et al. 2020), assumes an even more distant role, hidden behind the 'expert input' (Rieder and Skop 2021, p. 5) of challenge participants.

In sum, we argue that challenges are one of the most significant 'systems and lures' for effectively bringing together key AI assets. They serve as platforms where machine vision training datasets are employed for object detection, image segmentation, and motion prediction tasks, vital for autonomous vehicle development. By distributing and externalising specific AI tasks, challenges reduce the associated labour costs to nominal levels. Participating researchers are provided with the opportunity to tackle cutting-edge problems, gaining access to costly AI hardware otherwise out of reach (Luitse and Denkena 2021).

3 Taking up the challenge: a technographic approach

To investigate the role of the challenge as a structuring device in AI R&D, we adopt a material, 'technographic' approach (Bucher 2018; Van der Vlist et al. 2024). This aligns with the practices of digital STS (Vertesi and Ribes 2019) and involves gathering, analysing, and interpreting available information and materials from diverse sources to understand how applied R&D are organised and structured *around* Waymo.

Reflecting on ethnographic tactics for studying algorithmic systems, Seaver (2017, p. 6–7) emphasises the importance of 'glean[ing] information from diverse sources, even when … objects of study appear publicly accessible'. Ethnographers, like 'scavengers', piece together heterogeneous clues to gain partial insights into the complexities of the world. Adrian Mackenzie's (2017) ethnography of machine learners involved piecing together different aspects of the field of ML, from textbooks to statistical software packages such as R.

Steve Woolgar (1985), writing during the rise of 'expert systems', viewed AI work as an ongoing collaboration between human and machine actors, recognising the importance of studying 'the relationship between the pronouncements of spokesmen on behalf of AI and the *practical day-to-day activities* of AI researchers' (Woolgar 1985, p. 567, emphasis added). Overall, this perspective offers a fruitful avenue for investigating the role of challenges in the development of AI technologies: the 'many moments where explicit and implicit forms of human judgement come together with technical methods and artifacts' (Rieder and Skop 2021, p. 10).

Within the research literature on digital platforms, the diverse materials and documentation generated during such AI work are often referred to as 'boundary resources', serving the crucial function of facilitating and regulating the material aspects of participation for external third parties, extending beyond the platform itself (Van der Vlist 2022, p. 33). Critical scholars in media studies have explored how certain technical and informational resources, such as application programming interfaces (APIs) and reference documentation, shape power dynamics in various sectors of society, including digital marketing and advertising, mobile app development, and cultural production (Egliston and Carter 2022; Helmond et al. 2019; Helmond and Van der Vlist 2019; Ritala 2023). Drawing from these studies, our approach focuses on Waymo's pivotal role as a core platform company that provisionally brings together an autonomous vehicle technology ecosystem.

Waymo's Open Dataset Challenges are part of a larger collection of boundary resources that serve to 'convene' third-party developers and businesses, cultivating this ecosystem around Waymo. This 'convening' process, as described by Egliston and Carter, involves '"calling out to others, attracting their attention", requiring an "active response"' in the form of usage or participation (Egliston and Carter 2022, p. 10). In studying these various interactions and resources, we gain insights into the dynamics of collaboration and knowledge exchange that underpin Waymo's work.

This convening process undertaken by Waymo directly ties into the central argument of the challenge as an organising principle for AI. The challenge format serves as a crucial

mechanism through which Waymo brings together diverse stakeholders, including researchers, to collectively tackle cutting-edge autonomous vehicle problems. By convening participants through challenges, Waymo creates a platform for collaboration, competition, and knowledge sharing. The challenge format acts as an organising principle that shapes and directs the collective efforts of participants towards specific AI tasks and objectives. It serves as a focal point for mobilising expert labour, leveraging well-annotated training data, and harnessing the computational power necessary for advancing AI models and techniques. In this way, the challenge format not only facilitates the exploration of innovative solutions but also fosters the development of a vibrant ecosystem around Waymo's autonomous driving vision.

4 The setting: Waymo's Open Dataset and Challenges, 2019–2022

We embarked on our own 'scavenging-style' ethnography during a 3-day workshop held at the University of Siegen in late 2021. The primary objective of this workshop was to examine the Waymo Open Dataset, which served as our entry point into the study. Throughout the workshop, we immersed ourselves in the dataset by accessing the huge open datasets provided by Waymo. These files, each totalling 25 GB, encompassed various data types, structures, visual imagery, 3D models, and accompanying data attributes and image labels. In the process, we discovered a range of associated materials, documentation, and infrastructure dependencies linked to the Open Dataset, including the existence of annual Challenges.

Although we adapted Python scripts provided on Google Colaboratory (Colab) to facilitate the rendering of lidar images,[6] as we delved deeper into the dataset, we realised that we lacked the necessary ML skills to effectively work with it as intended. Consequently, we shifted from undertaking an exploratory (empirical) data project to an STS-oriented approach. At this stage, our engagement with the materials differed from that of a regular user or an empirical analyst. Instead, we assumed the role of 'scavengers', extracting insights from diverse sources and piecing together the available information to gain an understanding of the subject matter.

In the summer of 2022, we conducted a second workshop, entitled 'Taking up the Challenge'. During this workshop, our hands-on examination focussed on Waymo's Challenges and their connection to the broader 'research community' as defined by Waymo. Like the first, we collected and interacted with diverse materials and documentation available online, which provided valuable insights into these challenges. These materials included participant instructions, competition requirements, evaluative metrics, technical reports of ML models and methods, model output scores, challenge leaderboards, participant names, affiliated organisations, as well as research on previous challenge winners.

4.1 Workshop I: open dataset

In August 2019, Waymo introduced their Open Dataset initiative, announcing that they were 'sharing [their] self-driving data for research', and 'inviting the research community to join [them] with the release of the Waymo Open Dataset, a high-quality multimodal sensor dataset for autonomous driving' (Waymo 2019). It was described at the time as 'one of the largest, richest, and most diverse self-driving datasets ever released for research' (Waymo 2019).

The initial release of the Waymo Open Dataset consisted of data from 1000 'segments', with each segment capturing 20 s of continuous driving by Waymo autonomous vehicles. The primary focus was to provide 'researchers the opportunity to develop models to track and predict the behaviour of other road users' (Waymo 2019). The dataset encompassed data collected from various locations, including Phoenix (AZ), Kirkland (WA), Mountain View (CA), and San Francisco (CA) in the United States, capturing diverse environmental conditions such as 'day and night, dawn and dusk, sun and rain' (Waymo 2019). Each 20-s segment contained sensor data derived from five on-board lidar devices and five front-and-side-facing cameras. Notably, the dataset was extensively annotated, featuring 12 million 3D labels and 1.2 million 2D labels, playing a crucial role in training ML models for tracking and predicting the movement of vehicles in a driving environment.

In addition, the Open Dataset was available via Know Your Data (KYD),[7] a data exploration platform developed by Google, that 'helps researchers, engineers, product teams, and decision makers understand datasets with the goal of improving data quality, and helping mitigate fairness and bias issues' (Know Your Data 2023). By utilising KYD, users were able to navigate the contents of the dataset (of nearly a million items) and explore the relationships between various items. Furthermore, the images in the dataset were

---

[6] Colaboratory, or 'Colab', is a product developed by Google Research. It offers the capability for individuals to write and execute arbitrary Python code directly through a web browser. It is particularly tailored for tasks involving ML, data analysis, and educational purposes. In technical terms, Colab serves as a hosted Jupyter notebook service that boasts seamless usability without any initial setup requirements. In addition, it grants users free access to computational resources, including GPUs.

[7] 'waymo_open_dataset', Know Your Data, https://knowyourdata-tfds.withgoogle.com/#dataset=waymo_open_dataset.

labelled with Google Cloud Vision tags, providing additional information about road users ('TYPE_CYCLIST', 'TYPE_VEHICLE', 'TYPE_PEDESTRIAN', as well as 'has_faces', 'num_faces', etc.), various roadside objects ('Tree', 'Traffic light', 'Building', etc.), and other labels that could be utilised.

Waymo's open datasets, however, were not the first such datasets within the autonomous driving community. Waymo acknowledges the existence of the KITTI Vision Benchmark Suite, which was publicly released in March 2012, 7 years prior to Waymo's Open Datasets. The KITTI dataset is widely regarded as the benchmark for vision datasets in the field of autonomous driving and machine vision research.[8] Over the past decade, it has received updates, introduced novel benchmarks, and added newly annotated data. Given the popularity of existing benchmarks such as KITTI within the autonomous driving and machine vision communities, Waymo's decision to launch its open dataset naturally piqued our interest: what might they stand to gain from its release?

4.2 Workshop II: open dataset challenges

In March 2020, just as the COVID-19 pandemic was starting to impact Europe and the US, Waymo introduced their first Open Dataset Virtual Challenge. Waymo's principal scientist Drago Anguelov wrote that the newly-launched competition constituted 'the next phase of our program', with Waymo 'committed to fostering an environment of innovation and learning' (Anguelov 2020). The challenge comprised five specific machine vision challenges: 2D detection, 2D tracking, 3D detection, 3D tracking, and domain adaptation. Each challenge specified a task that participants were expected to perform with elements of the dataset, for example: 'given a set of camera images, produce a set of 2D boxes for the objects in the scene' or 'given a temporal sequence of lidar and camera data, produce a set of 3D upright boxes and the correspondences between boxes across frames' (Anguelov 2020).

Winners of each challenge were eligible for cash prizes, with $15,000 awarded to the first-place winners, $5000 for second place, and $2000 for third place. The competition opened on the same day as the announcement and ran until May 31, 2020. The leaderboard would be public and 'remain open for future submissions' (Anguelov 2020). Winners were also invited to present their winning methods at a workshop during the CVPR conference in Seattle, USA (Anguelov 2020). Subsequent editions of the Open Dataset

Challenge were announced in 2021 (Anguelov 2021) and 2022 (Waymo 2022a).[9]

For the 2021 edition, Waymo released a motion dataset for the first time, considered to be 'the largest interactive dataset yet released for research into behaviour prediction and motion forecasting for autonomous driving' (Anguelov 2021). The release included a comprehensive description of the datasets, a technical paper explaining the data annotation techniques used for the perception datasets (Qi et al. 2021), a Colab tutorial (Waymo 2021a), and a GitHub repository (Waymo 2022b). Four new challenges were introduced: motion prediction, interaction prediction, real-time 3D detection, and real-time 2D detection. The prize money remained the same, and participants were given a similar timeframe to submit their methods. Winners would once again be invited to present at the CVPR workshop, '[hoping] this expansion into motion data spurs on a new wave of research' (Anguelov 2021).

The 2022 edition followed a similar pattern, with the announcement in March, a submission deadline in May, and eligible winners presenting at the CVPR workshop in June. Waymo augmented the Open Dataset by adding additional labels to expand the range of tasks researchers could explore. These labels included 'key point labels' (capturing 'important small nuances'), '3D segmentation labels' (used to detect image pixels), and '2D-to-3D bounding box correspondence labels' ('to further enable research on sensor fusion of object detection and understanding') (Waymo 2022a). The challenges for 2022 included: motion prediction, occupancy and flow prediction, 3D semantic segmentation, and 3D camera-only detection.

During this second workshop, we decided to focus primarily on the detection challenges, offering a comparison of tasks, metrics, and methods across all three iterations (2020, 2021, 2022). The 2020 2D and 3D detection challenges evolved into real-time 2D/3D detection challenges in 2021 and further transformed into a 3D camera-only detection challenge in 2022. Participants had the opportunity to submit methods to previous challenges, enabling a temporal analysis of the original challenges (2D/3D detection) that laid the groundwork for these variations. To summarise, the 2021 edition introduced motion planning data for the first time, and in 2022, Waymo added additional labels to assist researchers in utilising the Open Dataset.

The materials and documentation encountered in relation to the Open Dataset Challenges originated from diverse sites and sources. These included the open dataset itself, the challenge guidelines, cloud computing tools and infrastructure,

---

[8] 'The KITTI Vision Benchmark Suite', KITTI, https://www.cvlibs.net/datasets/kitti/.

[9] A 2023 edition of the Waymo Open Dataset Challenge has subsequently concluded, with the workshop on autonomous driving (WAD) at CVPR running in June 2023, and reference to three sets of chal-

---

Footnote 9 (continued)

lenges organised by Waymo, Argo, and BDD (Berkeley DeepDrive), https://cvpr2023.wad.vision/.

and associated technical papers describing the submitted ML methods in greater detail. However, all resources were clearly related to the Open Dataset Challenges and served to convene the field of applied AI/ML research, engaging the research community in a manner that aligns with Waymo's business goals and strategy. Throughout, Waymo's parent company Google/Alphabet assumed a prominent role as the provider of cloud platform infrastructure (including computing resources and image labels from Google's Vision API), the host of the online code-sharing and notebook platform (Colab, linked to Google Drive), and the developer of the data exploration platform (KYD). Continuous discussions surrounding these materials and documentation took place during the workshops, forming the basis for further reflections in the article.

5 Shaping AI ecosystems through challenges: insights from Waymo's incremental approach

In the following, we detail six specific themes drawn from our study of Waymo's challenges: challenges as multifaceted *interfaces*, dynamics of *incrementalism*, the evolving significance of *metrics and benchmarks*, the *vernacular* of AI, the allure of *applied domains*, and the pursuit of *competitive advantages*. Collectively, these thematic insights provide a deeper understanding of challenges as central structuring devices that drive the advancement of autonomous vehicles and the broader realm of AI/ML. Within this context, challenges not only break down the intricate task of automating driving into feasible interim objectives but also serve as a manifestation of AI's operationalisation within specific domains or contexts. This operationalisation fuels inventive and exploratory endeavours evident in challenge submissions, where novel methods are trialled, traditional approaches serve as the foundation for innovation, and original combinations of data, algorithms, models, and workflows are tested, offering diverse pathways towards realising challenge goals.

5.1 Theme I: challenges as multifaceted interfaces

To begin with, Waymo's Open Dataset Challenges serve as conduits, or multifaceted *interfaces*, for a diverse array of components, including training datasets, annotations, leaderboards, arXiv and GitHub repositories, computer vision workshops (such as those hosted at CVPR), metrics, and methodologies. Collectively, these elements facilitate a dynamic interaction between autonomous vehicle companies and external stakeholders who harbour the potential to contribute significantly to the advancement of machine vision capabilities tailored for autonomous driving endeavours.

Even more so, challenges, ordinarily run on and through digital, and cloud-based platforms like those provided by Google, can be viewed as specialised 'interface methods' (Marres and Gerlitz 2015) in their own right, representing a convergence of diverse methodological traditions. In the current realm of scientific inquiry, collaboration transcends geographical boundaries, facilitated by an array of communication tools, diverse file formats, and an array of software analytical instruments, all seamlessly integrated through online platforms. Once-contained microchips have also broken free from the confines of laboratory equipment, finding expansive computational power across various cloud-based services, as elaborated here. As Vertesi and Ribes write, 'the textures of scientific and daily life at the beginning of the twenty-first century are suffused with online platforms and heterogenous informational environments' (Vertesi and Ribes 2019, p.1), of which the Waymo challenges are but one example.

However, the proliferation of computing possibilities, remote collaborators, disparate file formats, and analytical tools has augmented the need for robust organising principles, tangible mechanisms, and structural frameworks to ensure the seamless progression of AI technologies. This is precisely where the challenge format assumes its pivotal role, providing the very scaffolding required for challenge organisers to coalesce and harmonise disparate actors and endeavours. Paradoxically, in the challenges studied, it was the autonomous vehicle itself, as a comprehensive tangible entity, that receded from view, yielding to AI researchers' adherence to the 'logic of domains' (Ribes et al. 2019). Through this, the subject vehicle only necessarily returned in a spectral form, wholly 'decentered' (Law 2005, p. 32), as a mobile host for images captured by on-board sensing systems. Similarly, other vehicles materialised in spectral forms, reduced to clusters of dots and pixels that formed the bedrock for ML-oriented statistical inferences (Mackenzie 2017).

Nevertheless, orchestrating these endeavours typically falls to the most influential actors in the field like Waymo, with entities equipped to sway and direct researchers, engineers, start-ups, and even entire R&D divisions of companies to participate on terms carefully set by them. Whilst collaborations between academia and industry are not novel, especially within the world of AI/ML (Roland and Shiman 2002), the power dynamics between these entities have significantly evolved, with industry players now exercising greater control. These challenges in themselves challenge the conventional concept of 'competitions within the liberal order' (Stark 2020, p. 2) due to organisers' authority in shaping and configuring the terms and conditions of these competitions on an annual basis, aligning them intricately with the internal developmental trajectories of AI firms.

5.2 Theme II: dynamics of incrementalism

The Waymo challenges also exemplify a distinctive form of *incrementalism*, strategically designed to yield incremental improvements in object recognition and motion planning. A telling instance of this approach is found in the 2022 3D detection challenge, where a mere 0.018 difference separated the top-ranked method (0.7914 AP) and second place (0.7896 AP).[10] Over the course of the three years, the winning method in the same category rose from 0.7711 AP in 2020 to 0.7764 in 2021 to 0.7914 in 2022. Only in 2022 did any method post an AP score of over 0.79.[11] The significance of these seemingly marginal percentage gains becomes pronounced in the context of autonomous driving, as Srnicek (2022) contends. Such minute increments could indeed signify the distinction between a pedestrian or cyclist being struck, grazed, or entirely evaded by a vehicle. This trend might even be perceived as an extreme form of incrementalism, considering the quantitative subtlety (though qualitative importance) by which each successive winning method surpasses its predecessor. The specific definition of progress based on AP, encompassing all object categories, is of notable consequence, progressively elevating the performance threshold from year to year, persisting beyond the official challenge period. As Everingham et al. (2015, p. 133) observed regarding a previous object recognition challenge spanning 2005–2012, participants' optimal approach was to iteratively enhance the preceding year's winning method.

This ethos of incrementalism further manifested in the evolution of challenges, entailing refinements in task stipulations and parameters. Commencing in 2020, the 3D detection challenge solicited participants to generate a set of 3D upright boxes for scene objects (Waymo 2020), excluding any temporal component. In the subsequent year, Waymo introduced the real-time 3D detection challenge, retaining the original task specifications whilst introducing a temporal constraint (Waymo 2021b). The year 2022 witnessed the launch of a camera-only iteration of the challenge, restricting participants from incorporating lidar data into their methods (Waymo 2022c). With each iteration, challenge participants benefited from overarching enhancements and expansions to the foundational training dataset, encompassing a greater number of segments and an extended breadth of annotations.

Whilst all scientific and technical endeavours inherently encompass incremental progress, the broader concern emerges over whether these (highly) incremental gains are deemed sufficient by Big Tech firms financing the research and hinging their future growth on AI breakthroughs, particularly in areas like autonomous driving. This pertains equally to the broader public, who, in line with Reddy's proposition (1988), necessitate assurance that AI is delivering on its promises. Thinking critically, it is conceivable that these incremental advances might indeed reflect the sluggish, or potentially thwarted, efforts to realise automated driving witnessed in recent times (e.g. Korosec 2022). As argued by Everingham et al., the extreme incrementalism characteristic of such challenges poses the risk of 'reduc[ing] the diversity of methods within the community' as 'new methods that have the potential to give substantial improvements may be discarded before they have a chance to mature, because they do not yet beat existing mature methods' (Everingham et al. 2015, p. 133). In essence, the competitive structure fosters (extreme) incrementalism, as participants vie to surpass existing methods, thereby inhibiting the pursuit of what Everingham et al., (2015, p. 133) call methodological 'novelty'. Consequently, organised challenges crystallise a guiding ethos or value in the development of ML models, where prioritisation is accorded to 'a specific, quantitative, improvement over past work, according to some metric on a new or established dataset' (Birhane et al. 2022, p. 178).

5.3 Theme III: metrics and their evolving significance

Central to the orchestration of AI work within the Waymo challenges is the pivotal role of *metrics and benchmarks*, particularly Average Precision (AP), the preeminent standard for ML-based object recognition. The AP score plays an important role in these challenges, acting as the decisive arbiter for method validation. Any submission failing to attain a commendable AP score is categorically dismissed and invalidated. Consequently, the teams responsible for method design find themselves at a crossroads, necessitating a return to the proverbial drawing board, either to substantially refine and adapt their existing approach or devise an entirely new stratagem. Crucially, any such iteration must ultimately achieve a respectable AP score to merit consideration.

Nonetheless, the AP metric is neither arbitrary nor static; rather, it possesses a historical trajectory closely aligned with the timeline of the Waymo challenges themselves. The metrics employed by Waymo mirror the conventions established by the PASCAL Visual Object Classes (VOC) Challenge, conducted over 8 years (PASCAL VOC 2014; Everingham et al. 2015). In a pivotal decision, the organisers replaced the 'area under curve' (AUC) metric with AP

---

[10] Average precision or 'AP' is a statistical metric based on an 'intersection over union' (IOU) calculation, in which a decimal figure is generated based on the overlap of two areas. For example, a score of 1.00 AP would result from two areas overlapping entirely, and a score of 0.500 would result from two areas overlapping by exactly 50%.

[11] By November 19, 2022, the current front-runner had achieved an impressive score of 0.8132, whilst the methods in second and third place had also secured scores of 0.8 or above.

to enhance interpretability and other rationale (Everingham et al. 2010, p. 313). Notably, the introduction of the 3D camera-only detection challenge in 2022 led Waymo to introduce a modified version of AP termed LET-3D-APL, designed to accommodate 'depth estimation errors' (Hung et al. 2022, p. 1) common in monocular camera-based 3D object detection methods, inadequately addressed by the conventional AP metric. Metric transformations often stem from challenge organisers' insights into the shortcomings or intricacies of earlier metrics or recorded scores. Particularly intriguing is that the shift in the gold-standard metric was motivated by the transition from lidar-based to camera-based object detection, wherein the latter frequently presents longitudinal errors. Consequently, the 'LET' component in Waymo's new metric acronym, LET-3D-APL, stands for 'longitudinal error tolerance' (Hung et al. 2022), addressing a dimension that conventional AP failed to encompass but for which LET-3D-APL was meticulously crafted.

Metrics *matter* as they embody a 'golden rule' (Hind and Seitz 2022, p. 11), guiding engineers, computer scientists, and challenge participants alike: excellence is defined through superiority in the same problem domain relative to other methods (Birhane et al. 2022). However, challenge documentation and technical papers not only prescribe metrics but also underscore their social, contingent, and developmental nature. Just as methods are conceived, refined, modified, and adapted, metrics evolve in tandem. Echoing Everingham et al.'s perspective on the PASCAL VOC challenges, 'the metrics used in each … have typically been changed or refined at least once during the lifetime of the competition' (Everingham et al. 2015, p. 133). Tracing the evolutionary trajectory of metric establishment and implementation is paramount in comprehending the execution of machine vision tasks, fostering 'objectively and empirically measured' performance so 'the community [can] know what really work[s]' (Zissermann et al. 2012, p. 2082). This is of particular import since changes in key metrics might inadvertently present a parallel quandary to the broader challenge framework: 'methods may become overly tailored to the specific evaluation metrics chosen for each competition' (Everingham et al. 2015, p. 133). In essence, such challenges may inadvertently narrow the scope of competition (Stark 2020), with metrics fundamentally instrumental in shaping this dynamic. What sets apart this 'metric work' from other instances of rule adjustments is its inherent recognition of partiality and flexibility, invariably rendering metrics 'adequate' for the task at hand.

5.4 Theme IV: vernacular of AI work

Within the realm of the Waymo challenges lies a nuanced and intricate *vernacular* of AI work, characterised by two distinct dimensions. Firstly, a pronounced element of playfulness permeates the nomenclature, resonating with colloquial phrases that infuse the machine vision landscape. This tendency is evident in the technical papers and GitHub repositories, where terms like 'bells and whistles' are used to denote methods without additional embellishments or complex features (Bergman et al. 2019; Liu et al. 2022; Yin 2021; Zhang et al. 2020). We also found references to other established, off-the-shelf ML models with memorable (and meme-able) names, like 'YOLO' (Redmon et al. 2016). Rather than 'You Only Live Once', it stands for 'You Only *Look* Once', referring to the total snapshot the model makes of test images, instead of multiple passes common to other models.[12] This playful linguistic aspect can be seen as an extension of practices found within the 'hacker class' (Wark 2004), drawing inspiration from social media and meme culture (Dal Dosso et al. 2021). As Gabriella Coleman notes, 'hackers … ha[ve] an exhaustive ability to "misuse" most anything and turn it into grist for the humor mill' (Coleman 2013, p. 7). Once one is able to 'master the esoteric and technical language' of the work being performed, 'a rich terrain of jokes bec[o]me[s] sensible' (Coleman 2013, p. 7), something eminently discernible in the Waymo challenges.

Second, this vernacular exhibits a more structured facet, often utilising physiological and neurological metaphors to elucidate the composition of methods themselves. The names, descriptions, and accompanying technical documents of these methods are replete with expressions and allusions to fundamental components and the lineage of prior work that underpins each submission. Method names, for instance, typically refer to core elements and supplementary features. PV-RCNN++, a team in the 3D object-detection challenge (Shi 2022), signifies a 'Point Voxel-RCNN' architecture with additional appended features ('++'). Similarly, 'CenterTrans_V3', another submission, denotes the third iteration of a fusion between two methods, CenterPoint and Transformer (Zhang 2022). Whilst familiarity with fundamental components (such as RCNNs) eases the interpretation of these names, they generally adhere to a comprehensible structure.

Whilst ML itself is steeped in neurological metaphors, exemplified by the notion of 'neural' networks, teams routinely refer to 'backbone', 'neck', and 'head' segments of their models [Fig. 1]. For instance, the PV-RCNN++ method incorporates a '3D voxel CNN with sparse convolution [used] as the backbone' (Shi et al. 2021, p. 3), whilst the 'TS-LidarDet' submission emphasises the role of 'necks' in providing 'interfaces to build complementary feature extraction layers' (Chen 2020). Another method, 'BEVFusion-TTA' introduces the concept of 'task-specific heads' (Liu et al. 2022, p. 3) for distinct detection tasks.
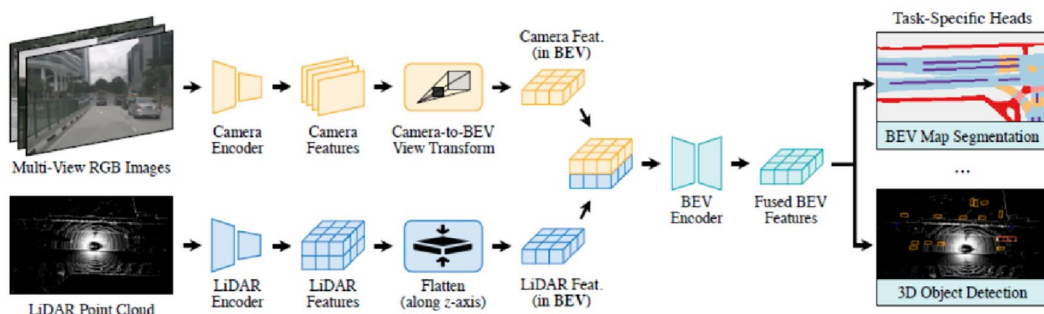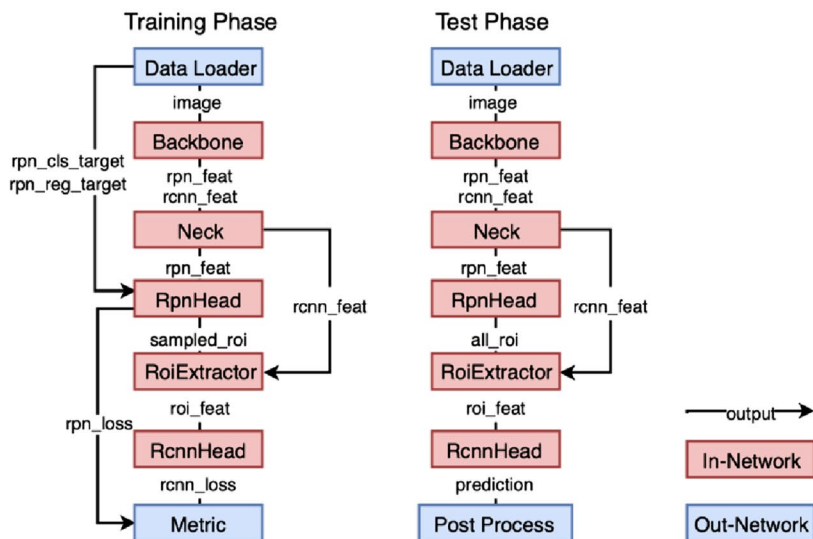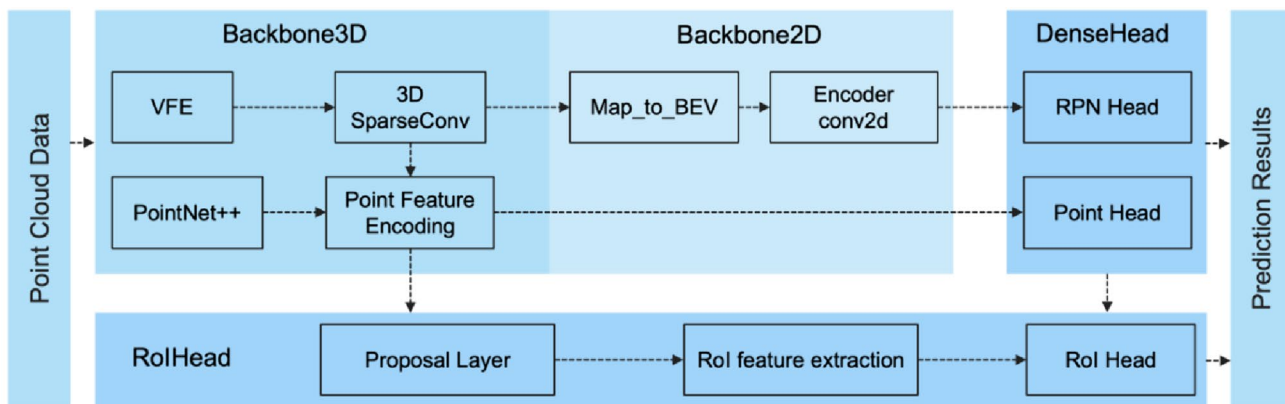
---

[12] https://pjreddie.com/darknet/yolo/.

Fig. 1 Three ML model frameworks showcasing the concepts of 'backbones', 'necks', and 'heads'. Sources: OpenPCDet (2020), Simpledet (2019), and Liu et al. (2022, p. 3)

This descriptive vernacular presents a revelation of considerable logical and overarching significance: the neural terminology extends beyond cognitive metaphors to comprehensively encompass the human nervous system. As Mackenzie (2017, p. 182) suggests, the portrayal of neural networks occasionally shifts towards human subjects or, more specifically, human brains, whilst at other junctures, it assumes an informational essence. In this landscape, the brain-centred metaphors recede, replaced by an expressive neurological lexicon. This structural metaphor empowers the community with an accessible framework for comparative comprehension. It signifies the 'core' elements of a model (the backbone) likely for off-the-shelf deployment, the linkages connecting them (the neck), and the specialised

components tailored for distinct tasks and contexts (the head). In this descriptive juncture, the models vividly and conspicuously convey their modular form and functionality.

5.5 Theme V: allure of the applied domain

The allure of extensive, diverse, and meticulously annotated training data reverberated strongly amongst potential participants. Waymo's cognisance of this inherent appeal was underscored by their rationale for initiating the challenges (Anguelov 2020), with the intention of elevating the Waymo Open Dataset to a preeminent training data benchmark. This strategic move aimed to position the dataset in competition with, or even as a potential successor to, established datasets like KITTI (Geiger et al. 2012) or nuScenes (Caesar et al. 2020), known for their significance in advancing machine vision research.

For participants, the allure was also related to the prospect of applying their burgeoning machine vision skills, a significant number of whom were either computer science PhD students or recent graduates. The challenges provided a unique avenue to not only exercise their newly acquired general skills on well-defined (machine vision) problems within real-world contexts but also to engage with trendy or cutting-edge *domains* such as autonomous driving. By affording participants the opportunity to channel their expertise in such a way, Waymo harnessed the synergy between emergent skills and pertinent application.

The Waymo challenges granted participants access to sensor data and computational resources that would otherwise be unreachable or unfeasible, inaccessible to most researchers unless affiliated with esteemed institutions like CMU or Stanford. Whilst direct participant interviews were not conducted, the broad representation across diverse locations and institutions strongly implies the pronounced allure of these factors. This confirms Srnicek's assertion that AI start-ups 'remain dependent on AI providers' for computational infrastructure (Srnicek 2022, p. 244). For Waymo, these challenges serve as a conduit to gather real-world data, establishing a foundation based on predetermined computational benchmarks. For instance, the 2021 2D Detection challenge demanded submission latency measurements on a specific cloud-based tensor core GPU, the Nvidia Tesla v100 GPU, one of Google's seven GPU platforms dedicated to cloud-based ML training (Google 2023a). Tying competition requirements to available computational resources naturally enables Waymo to scale their operations, ultimately a considerable and ongoing concern (Sharp and Pan 2022).

Surprisingly, a notable proportion of participants were from Chinese universities, research labs, and AI start-ups. This was an unexpected finding, despite—or perhaps because of—China's huge investment in AI (Lucas 2017). This make-up diverged starkly from the participant composition of earlier events like the DARPA Grand Challenge, which primarily featured US-based teams from institutions like Stanford, CMU, and Virginia Tech. Podium teams in the 2022 3D camera-only detection challenges were affiliated to Chinese institutions, including the Shanghai AI Lab and the Chinese University of Hong Kong, with additional affiliations to the Mohamed bin Zayad University of Artificial Intelligence (UAE), and Pegasus Tech, a Silicon Valley-based venture capital firm.

Whilst participants from US institutions, including MIT, remained prominent, the collective participant landscape clearly indicated a multipolar evolution in AI R&D. This shift from a US-centric paradigm involves different kinds of commercial actors (DiDi Global, Horizon Robotics, etc.), whilst being significantly led by Chinese-affiliated institutions and actors. This departure also marks a substantial deviation from past eras such as the 2000s DARPA Grand Challenge, and the 1980s–1990s machine intelligence period, both of which were US state-led initiatives, involving commercial partnerships with US firms. This transformative shift, as evident through the dynamics of the Waymo challenges, provides a snapshot of machine vision's trajectory within the intense capital influx into AI/ML technologies between 2020 and 2022.

However, this era of collaboration might potentially be nearing its peak, following export controls imposed by the US government targeting technology flow to China (Sevastopulo and Hille 2022). This policy, with a specific focus on the semiconductor and AI sectors, has introduced an aura of apprehension akin to historical instances like Japan's Fifth Generation program in the late 1980s (Roland and Shiman 2002, p. 2). In short, the current landscape of burgeoning competition and collaboration in AI, as convened by Waymo's challenges, might represent a pivotal moment where US stakeholders mirror historical concerns about emerging technological rivals, or at the very least, it indicates the geopolitical and political–economic stakes involved in shaping AI development.

5.6 Theme VI: securing competitive advantage

The Waymo challenges serve as a strategic embodiment of a well-established Big Tech R&D playbook, aiming to shape and 'lock-in' (Urry 2004) a thriving developer community within their prescribed timelines, developmental trajectories, and technical frameworks. This concerted effort brings young researchers into Google's orbit, offering tools and services like Google Colab and Tensorflow in exchange for labour, extending how Google uses its online Machine Learning Crash Course (MLCC) programme to hook users in the first place (Luchs et al. 2023). The challenge format itself stands as a relatively tried-and-tested format for achieving this goal, enabling both organisers and entrants 'to

commit time and funds to the competition' (Kreiner 2020, p. 51) in an efficient, compact manner.

Cost-effectiveness underpins this strategic approach, as running a challenge for external participants, coupled with a modest cash prize fund ($15,000 for winners), proves economically advantageous compared to hiring full-time engineers at market rates.[13] Whilst the expenses associated with constructing training datasets and scaling computational resources are substantial, they are spread across the broader operations of Google/Alphabet, as well as being specifically valuable to Waymo's own internal initiatives. The competitive spirit engendered by the challenges, coupled with the incentive of the prize fund, acts as a powerful catalyst propelling participants to invest substantial time and energy into the intricate tasks of method development, rigorous testing, meticulous verification, comprehensive documentation, and final submission. Vertesi et al. (2021) aptly term this phenomenon the 'pre-automation' phase of AI, characterised by companies' rapid AI product scaling through the internalisation of highly skilled technical endeavours.

In contrast to the veiled realm of 'temporary, vendor, and contractor' (TVCs) workers, used by Big Tech firms to plug gaps in short- and mid-term product development (Brophy and Grayer 2021), the temporary labour demonstrated by challenge participants is openly documented and is even celebrated as a rite of passage. This holds especially true for numerous aspiring young computer scientists who eagerly embrace the opportunity to apply their newfound knowledge to cutting-edge challenges. An indicator of this recognition and pride can be found in the frequent referencing of podium achievements on the GitHub pages of participating teams, where these accolades are displayed as prestigious badges of honour (e.g. BEVFormer 2023). Here, if Google's MLCC programme allows them to recruit 'new AI talent' (Luchs et al. 2023, p. 9) at one end of the AI talent 'pipeline', Waymo's challenges offer the opportunity to channel and *celebrate* that talent at the other.

In the 2023 edition, only challenge winners receive a prize, capped at $10,000 in Google Cloud credits (Anguelov 2023). Using their own Cloud Pricing Calculator (Google 2023b), $10,000 would offer a team roughly 3 month's access to Google's second-generation (v2) Cloud TPU

service, useful for training ML models remotely.[14][15] Steinhoff (2023), building on Rikap (2021), characterises this phenomenon as a 'subordinated innovation network', further devaluing the work of those competing in such ML challenges (Steinhoff 2022). The shift from hard cash to credits further entrenches this subordination, intensifying participant dependence, whilst hardening the resultant innovation network.

Learning lessons from the PASCAL VOC challenges, Everingham et al. (2015) suggested that the open format tended to reduce the diversity of methods within the wider research community. If participants wanted to win, they stood the best chance by making 'an incremental improvement on the previous year's winning method' (Everingham et al. 2015, p. 133), rather than develop new methods from scratch. In the context of the Waymo challenges, this predilection locks participants into Google products for model training, strengthening the gravitational pull towards the Google/Alphabet ecosystem. Everingham et al. (2015, p. 132) also remarked that having software able to 'run everything "out of the box"', from training to validation was crucial. In this respect, Waymo goes a step further by being both software developer and challenge host organiser, a unique position that asserts their monopoly power. Paradoxically, such a situation could hamper broader progress, following Everingham et al. (2015), diverting attention away from maturing methods and fostering an environment of heightened incrementalism.

Waymo is not the only firm to run an AI challenge within the autonomous vehicle domain. However, the decision by Ford to shutter Argo AI (Korosec 2022) has arguably diluted the prominence and impact of their rival Argoverse initiative,[16] now lacking the envisaged pipeline from challenge participation to commercial deployment. Essentially, Waymo has solidified a monopoly position within AI challenges, bolstering their *competitive advantage* by maintaining their sustained presence.

6 Conclusion: conceptualising challenges as an organising principle in AI innovation

Throughout this article, we have explored the fundamental role of challenges in shaping AI development. By juxtaposing the era of Grand Challenges with Waymo's strategy of incremental challenges within the realm of autonomous

---

[13] As of August 24, 2023, a Waymo software engineer can expect to receive a base pay of $135,679, https://www.glassdoor.com/Salary/Waymo-Software-Engineer-Salaries-E1635890_D_KO6,23.htm.

[14] Considering the utilisation of 4 TPU chips (8 nodes) on a single device, situated in Iowa (USA) and running non-stop for 24 h a day, the projected cost would amount to $3285 per month. This estimation is derived from calculations made on May 18, 2023, and reflects the cumulative expense over a three-month period, totalling $9855. See: 'Google Cloud Pricing Calculator', https://cloud.google.com/products/calculator/#id=1b1037e8-4c8f-4c75-bfc2-a6e6d7e5d9d5.

[15] Google's TPU Research Cloud (TRC) program also offers researchers access to cloud-based TPUs at monetary cost, although participants must share any research resulting from TPU access publicly, https://sites.research.google/trc/about/.

[16] 'Argoverse', https://www.argoverse.org/.

driving, we have unveiled a prevailing approach that characterises both Waymo and the broader contemporary AI landscape. Our investigation of the specific objects and practices of researchers in this field contributes to the critical literature in STS and media studies, shedding light on the history of AI research in the self-driving industry and the history of challenges within this area. In addition, it highlights the ongoing significance of the infrastructures supporting this research, as open datasets and challenges emerge as crucial instruments shaping research funding and the political economy of AI. Despite the existence of alternative paths and resistance to incrementalism, Waymo's challenge initiative has effectively provided a platform for scaling R&D efforts whilst engaging external participants who share their interests. Through the provision of quality training data and computational resources, Waymo has cultivated a global research community united by common objectives, set by themselves.

Despite its scope, our exploration has merely grazed the surface of Waymo's multifaceted efforts in shaping the contours of AI R&D, and there are several avenues that warrant further inquiry. These might be categorised according to the scale of investigation: challenges as practices, challenges as economic phenomenon and challenges as instances of the infrastructuralisation of AI/ML. In the first instance, examining the distribution of machine vision labour, including the division of tasks within challenge teams, could yield profound insights. Exploring team methodologies, organisational structures, workflow plans, and the strategic leveraging of prior work are pivotal for understanding the nuanced costs and benefits encountered by potential challenge participants. A more sustained focus on the role of ML and machine vision metrics—how they are devised, who designs them, and what they replace—might also shed some light on the contingencies and power dynamics of ML practices writ large.

Likewise, comparing different challenges, challenge formats, and challenge platforms would offer an insight into cross-domain, cross-format, and cross-platform themes. Luchs et al.'s (2023) comparison between online ML courses run by Google and IBM, for example, suggests divergent approaches to offering practical ML experience to computer/data scientists. Steinhoff (2022; 2023) and Rikap's (2021) work also point towards the possibility of evaluating how ML and 'data science work' (Steinhoff 2022, p. 193) conducted for such challenges is being shaped by automation, evidencing how AI/ML firms seek to reduce the huge financial costs for building ML models, products, and platforms. In other words, Waymo's own challenges are not necessarily unique, but provide evidence of a certain challenge 'playbook' to be found across different AI/ML domains.

As the focus shifts from autonomous driving to the hype around large-language models (LLMs), the importance of machine vision challenges, including those hosted by Waymo and its competitors, may undergo changes for aspiring computer scientists. It remains to be seen when Waymo might reassess its developmental roadmap and evaluate the sustainability and usefulness of organising external challenges. The transition from cash prizes to Google Cloud credits indicates a potential shift in priorities, aiming to consolidate and optimise investments in autonomous driving. The recent suspension of '24/7' Cruise passenger services in San Francisco suggest the autonomous vehicle battle has already entered another stage of development altogether (Hawkins 2023), despite Waymo expanding operations to Los Angeles (Davis 2024).

Beyond the specific trajectory of Waymo and the autonomous driving applied domain, further research is necessary to explore the broader infrastructuralisation and *industrialisation* of AI (Van der Vlist et al. 2024). Central to this exploration is understanding the commodification of LLMs and the widespread proliferation of third-party services that pivot on models like ChatGPT. Scrutinising intricate relationships, evolving licensing models, and the emergence of counter-LLM platforms across diverse sectors, such as higher education, to monitor LLM-generated content, presents compelling questions that stretch far beyond this article's scope but necessitate concerted attention in forthcoming research.

In conclusion, this article underscores the broader historical and critical importance of challenges as a pivotal organising principle shaping AI development, with Waymo's incremental approach serving as a prominent example in the field today. By investigating the dynamics and characteristics of these challenges, and their materiality and infrastructures, scholars can gain valuable insights into the trajectory of AI development and its driving forces in specific industry sectors like self-driving technology. However, the analysis also extends beyond this to the wider AI/ML landscape, offering a nuanced understanding of how challenges shape the contours of technological progress and influence broader socioeconomic trends. As such, scholars can leverage challenges as an entry point into AI research, using them as a lens to critically examine the interplay between technological innovation, industry dynamics, and societal change.

Curmudgeon Corner Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby

---

===

Hind S (2023) Sensor work: enabling the inter-operation of autonomous vehicles. J Comput-Mediat Commun 28(5):1–10. https://doi.org/10.1093/jcmc/zmad014

Hind S, Gekker A (2022) Automotive parasitism: examining Mobileye's 'car-agnostic' platformisation. New Media Soc. https://doi.org/10.1177/14614448221104209

Hind S, Seitz T (2022) Cynical technical practice: from AI and APIs. Convergence. https://doi.org/10.1177/13548565221133248

Hind S, Kanderske M, van der Vlist FN (2022) Making the car 'platform ready': how Big Tech is driving the platformisation of automobility. Soc Media Soc 8(2):1–13. https://doi.org/10.1177/20563051221098697

Hooper J (2004) From Darpa Grand Challenge 2004: DARPA's debacle in the desert. Popular Science. https://www.popsci.com/scitech/article/2004-06/darpa-grand-challenge-2004darpas-debacle-desert/. Accessed 21 Dec 2023

Hung W-C, Kretzschmar H, Casser V, Hwang J-J, Anguelov D (2022) LET-3D-AP: longitudinal error tolerant 3D average precision for camera-only 3D detection. arXiv. https://doi.org/10.48550/arXiv.2206.07705

Iapaolo F (2023) The system of autono-mobility: computer vision and urban complexity—reflections on artificial intelligence at urban scale. AI Soc 38(3):1111–1122. https://doi.org/10.1007/s00146-022-01590-0

Know Your Data (2023) Know Your Data. Google. https://knowyourdata.withgoogle.com/. Accessed 21 Dec 2023

Korosec K (2022) Ford, VW-backed Argo AI is shutting down. TechCrunch. https://techcrunch.com/2022/10/26/ford-vw-backed-argo-ai-is-shutting-down/. Accessed 21 Dec 2023

Kreiner K (2020) Pick the winner, so you can then choose the reasons: epistemic dissonance in architecture competitions. In: Stark D (ed) The performance complex: competition and competitions in social life. Oxford University Press, Oxford, pp 31–54

Law J (2005) Aircraft stories: decentering the object in technoscience. Duke University Press, Durham

Liu Z., Tang H, Amini A, Yang X, Mao H, Rus D, Han S (2022) BEVFusion: Multi-task multi-sensor fusion with unified bird's-eye view representation. Preprint at https://arxiv.org/abs/2205.13542. Accessed 21 Dec 2023

Lucas L (2017) China seeks dominance of global AI industry. Financial Times. https://www.ft.com/content/856753d6-8d31-11e7-a352-e46f43c5825d. Accessed 21 Dec 2023

Luchs I, Apprich C, Broersma M (2023) Learning machine learning: on the political economy of big tech's online AI courses. Big Data Soc 10(1):1–12. https://doi.org/10.1177/20539517231153806

Luitse D, Denkena W (2021) The great transformer: examining the role of large language models in the political economy of AI. Big Data Soc 8(2):1–14. https://doi.org/10.1177/20539517211047734

Mackenzie A (2017) Machine learners: archaeology of a data practice. MIT Press, Cambridge

Markoff J (2010) Google cars drive themselves, in traffic. The New York Times. https://www.nytimes.com/2010/10/10/science/10google.html. Accessed 21 Dec 2023

Marres N (2020) Co-existence or displacement: do street trials of intelligent vehicles test society? Br J Sociol 71(3):537–555. https://doi.org/10.1111/1468-4446.12730

Marres N, Gerlitz C (2015) Interface methods: renegotiating relations between digital social research. STS Sociol Sociol Rev 64(1):21–46. https://doi.org/10.1111/1467-954X.12314

OpenPCDet (2020) Model framework. Github. https://github.com/open-mmlab/OpenPCDet/blob/master/docs/model_framework.png. Accessed 21 Dec 2023

Pan A., Fields R (2022) D. Sculley—Technical Debt, Trade-offs, and Kaggle. Gradient Dissent. https://wandb.ai/wandb_fc/gradient-dissent/reports/D-Sculley-Technical-Debt-Trade-offs-and-Kaggle--VmlldzozMDU3NDU0. Accessed 21 Dec 2023

PASCAL VOC (2014) The PASCAL Visual Object Classes homepage. PASCAL VOC. http://host.robots.ox.ac.uk/pascal/VOC/. Accessed 21 Dec 2023

Qi CR, Zhou Y, Najibi M, Sun P, Vo K, Deng B, Anguelov D (2021) Offboard 3D object detection from point cloud sequences. arXiv. https://doi.org/10.48550/arXiv.2103.05073

Reddy R (1988) Foundations and grand challenges of artificial intelligence. AI Magaz 9(4):9–21

Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. arXiv. https://doi.org/10.48550/arXiv.1506.02640

Rella L (2023) Close to the metal: towards a material political economy of the epistemology of computation. Soc Stud Sci. https://doi.org/10.1177/03063127231185095

Ribes D (2019) STS, meet data science, once again. Sci Technol Human Values 44(3):514–539. https://doi.org/10.1177/0162243918798899

Ribes D, Hoffman AS, Slota SC, Bowker GC (2019) The logic of domains. Soc Stud Sci 49(3):281–309. https://doi.org/10.1177/0306312719849709

Rieder B, Skop Y (2021) The fabrics of machine moderation: studying the technical, normative, and organizational structure of Perspective AI. Big Data Soc 8(2):1–16. https://doi.org/10.1177/20539517211046181

Rikap C (2021) Capitalism, power and innovation: intellectual monopoly capitalism uncovered. Routledge, London

Ritala P (2023) Grand challenges and platform ecosystems: scaling solutions for wicked ecological and societal problems. J Prod Innov Manag. https://doi.org/10.1111/jpim.12682

Roland A, Shiman P (2002) Strategic computing: DARPA and the quest for machine intelligence, 1983–1993. MIT Press, Cambridge

Sánchez-Monedero J, Dencik L (2022) The politics of deceptive borders: 'biomarkers of deceit' and the case of iBorderCtrl. Inf Commun Soc 25(3):413–430. https://doi.org/10.1080/1369118X.2020.1792530

Seaver N (2017) Algorithms as culture: some tactics for the ethnography of algorithmic systems. Big Data Soc 4(2):1–12. https://doi.org/10.1177/2053951717738104

Sevastopulo D, Hille K (2022) US hits China with sweeping tech export controls. Financial Times. https://www.ft.com/content/6825bee4-52a7-4c86-b1aa-31c100708c3e. Accessed 21 Dec 2023

Sharp C, Pan A (2022) Drago Anguelov—Robustness, safety, and scalability at Waymo. Weights & Biases. https://wandb.ai/wandb_fc/gradient-dissent/reports/Drago-Anguelov-Robustness-Safety-and-Scalability-at-Waymo--VmlldzoyMzAwNzY1. Accessed 21 Dec 2023

Shi S (2022) PV-RCNN++. Waymo. https://waymo.com/open/challenges/entry/?challenge=DETECTION_3D&challengeId=DETECTION_3D&emailId=82494e86-e86d&timestamp=1648858093038803. Accessed 21 Dec 2023

Shi S, Guo C, Jiang L, Wang Z, Shi J, Wang X, Li H (2021) PV-RCNN: Point-voxel feature set abstraction for 3D object detection. arXiv. https://arxiv.org/abs/1912.13192. Accessed 21 Dec 2023

Simpledet (2019) Detector. Github. https://github.com/open-mmlab/OpenPCDet/blob/master/docs/model_framework.png. Accessed 21 Dec 2023

Sprenger F (2022) Microdecisions and autonomy in self-driving cars: virtual probabilities. AI Soc 37:619–634. https://doi.org/10.1007/s00146-020-01115-7

Srnicek N (2022) Data, compute, labor. In: Graham M, Ferrari F (eds) Digital work in the planetary market. MIT Press, Cambridge, pp 241–261

Stark D (2020) The performance complex. In: Stark D (ed) The performance complex: competition and competitions in social life. Oxford University Press, Oxford, pp 1–27

Steinberg M (2022) From automobile capitalism to platform capitalism: Toyotism as a prehistory of digital platforms. Organization Studies 43(7):1069–1090. https://doi.org/10.1177/01708406211030681

Steinhoff J (2022) The proletarianization of data science. In: Graham M, Ferrari F (eds) Digital work in the planetary market. MIT Press, Cambridge, pp 191–206

Steinhoff J (2023) AI ethics as subordinated innovation network. AI Soc. https://doi.org/10.1007/s00146-023-01658-5

Stilgoe J (2017) Machine learning, social learning and the governance of self-driving cars. Soc Stud Sci 48(1):26–56. https://doi.org/10.1177/0306312717741687

Tennant C, Stilgoe J (2021) The attachments of 'autonomous' vehicles. Soc Stud Sci 51(6):846–870. https://doi.org/10.1177/03063127211038752

Thrun S, Montemerlo M, Dahlkamp H, Stavens D, Aron A, Diebel J, Fong P, Gale J, Halpenny M, Hoffmann G, Lau K, Oakley C, Palatucci M, Pratt V, Stang P, Strohband S, Dupont C, Jendrossek L-E, Koelen C, Markey C, Rummel C, Jv N, Jensen E, Alessandrini P, Badski G, Davies B, Ettinger S, Kaehler A, Nefian A, Mahoney P (2007) Stanley: the robot that won the DARPA Grand challenge. In: Buehler M, Iagnemma K, Singh S (eds) The 2005 DARPA grand challenge: the great robot race. Springer, New York, pp 1–43

Thylstrup NB (2022) The ethics and politics of data sets in the age of machine learning: deleting traces and encountering remains. Media Cult Soc 44(4):655–671. https://doi.org/10.1177/01634437211060226

Tubaro P, Casilli AA, Coville M (2020) The trainer, the verifier, the imitator: three ways in which human platform workers support artificial intelligence. Big Data Soc 7(1):1–12. https://doi.org/10.1177/2053951720919776

Urry J (2004) The 'system' of automobility. Theory Cult Soc 21(4/5):25–39. https://doi.org/10.1177/0263276404046059

Vertesi J, Ribes D (2019) Introduction. In: Vertesi J, Ribes D (eds) DigitalSTS: a field guide for science and technology studies. Princeton University Press, Princeton NJ, pp 1–10

Vertesi JA, Goldstein A, Enriquez D, Liu L, Miller KT (2021) Pre-automation: insourcing and automating the gig economy. Sociologica 14(3):167–193. https://doi.org/10.6092/issn.1971-8853/11657

van der Vlist FN (2022) The platform as ecosystem: configurations and dynamics of governance and power. Utrecht University, Utrecht. https://doi.org/10.33540/1284

van der Vlist FN, Helmond A, Ferrari FL (2024) Big AI: cloud infrastructure and the industrialisation of artificial intelligence. Big Data Soc 11(1):1–16. https://doi.org/10.1177/20539517241232630

Wark M (2004) A Hacker Manifesto. Harvard University Press, Cambridge

Waymo (2019) Waymo Open Dataset: Sharing our self-driving data for research. Waymo. https://blog.waymo.com/2019/08/waymo-open-dataset-sharing-our-self.html. Accessed 21 Dec 2023

Waymo (2020) Challenge: 3D detection. Waymo. https://waymo.com/open/challenges/2020/3d-detection/. Accessed 15 Apr 2024

Waymo (2021a) Waymo Open Dataset motion tutorial. Waymo. https://blog.waymo.com/2022/03/expanding-waymo-open-dataset-with-new-labels.html. Accessed 21 Dec 2023

Waymo (2021b) Challenge 3: Real-time 3D detection. Waymo.https://waymo.com/open/challenges/2021/real-time-3d-prediction/. Accessed 21 Dec 2023

Waymo (2022a) Expanding the Waymo Open Dataset with new labels and challenges. Waymo.https://blog.waymo.com/2022/03/expanding-waymo-open-dataset-with-new-labels.html. Accessed 21 Dec 2023

Waymo (2022b) Waymo Open Dataset. Waymo. https://github.com/waymo-research/waymo-open-dataset. Accessed 21 Dec 2023

Waymo (2022c) Challenge: 3D camera-only detection. Waymo.https://waymo.com/open/challenges/2022/3d-camera-only-detection/. Accessed 21 Dec 2023

Waymo (2023) Safety. Waymo. https://waymo.com/safety/. Accessed 21 Dec 2023

Woolgar S (1985) Why not a sociology of machines? The case of sociology and artificial intelligence. Sociology 19(4):557–572. https://doi.org/10.1177/0038038585019004005

Yin T (2021) CenterPoint. Github. https://github.com/tianweiy/CenterPoint. Accessed 21 Dec 2023

Zhang X (2022) CenterTrans_V3. Waymo. https://waymo.com/open/challenges/entry/?challenge=DETECTION_3D&challengeId=DETECTION_3D&emailId=90669e6a-4c25&timestamp=1653461849479974. Accessed 21 Dec 2023

Zhang S, Chi C, Yao Y, Lei Z, Li SZ (2020) Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. arXiv. https://doi.org/10.48550/arXiv.1912.02424

Zisserman A, Winn J, Fitzgibbon A, Van Gool L, Sivic J, Williams C, Hogg D (2012) In memoriam: Mark Everingham. IEEE Trans Pattern Anal Mach Intell 34(11):2081–2082. https://doi.org/10.1109/TPAMI.2012.204