a survey of large language models

A Survey of Large Language Models: Exploring the Giants of AI

a survey of large language models reveals a fascinating landscape where artificial intelligence is reshaping how we interact with technology, communicate, and even create content. These models, powered by deep learning, have grown exponentially in size and complexity over the past few years, enabling machines to understand and generate human-like text with astonishing accuracy. Whether you're a tech enthusiast, a developer, or simply curious about AI, understanding these behemoths is essential to grasp the future trajectory of language technology.

Understanding Large Language Models

Large language models (LLMs) are neural networks trained on massive amounts of text data to perform a variety of natural language processing (NLP) tasks. They generate text, translate languages, summarize information, answer questions, and even engage in creative writing. But what exactly makes these models "large," and why does size matter?

The Scale and Architecture

The term "large" typically refers to the number of parameters within the model — essentially, the knobs and dials the model adjusts during training to learn patterns in language. Early language models had millions of parameters, but today's giants boast billions or even trillions. For example, OpenAI's GPT-3 contains 175 billion parameters, while newer models continue to push this boundary.

This scale allows LLMs to capture subtle nuances in language, context, and meaning that smaller models might miss. Architecturally, most modern LLMs use the Transformer architecture, known for its ability to process sequences of data efficiently and capture long-range dependencies in text.

Why Size Isn't Everything

While increasing parameters often improves performance, it also brings challenges like higher computational costs, longer training times, and environmental concerns due to energy consumption. Moreover, beyond a certain point, simply adding parameters yields diminishing returns without improvements in training data quality or model design.

This is why researchers are exploring techniques such as fine-tuning, knowledge distillation, and prompt engineering to make large models more efficient and adaptable without necessarily increasing size.

A Survey of Prominent Large Language Models

Taking a closer look at some of the most influential large language models helps illustrate the diversity and capabilities currently available.

GPT Series by OpenAI

The Generative Pre-trained Transformer (GPT) series arguably popularized large language models. Starting with GPT, then GPT-2, and the breakthrough GPT-3, these models showcased impressive abilities in text generation, coding assistance, and conversational AI.

- **GPT-2:** Released in 2019 with 1.5 billion parameters, it surprised many with its ability to generate coherent paragraphs and stories.
- **GPT-3:** The leap to 175 billion parameters significantly improved performance across various NLP benchmarks and made it a versatile tool for applications ranging from chatbots to creative writing.

OpenAI also introduced ChatGPT, a fine-tuned version optimized for dialogue, which gained massive popularity for its human-like conversational skills.

BERT and Its Variants

Bidirectional Encoder Representations from Transformers (BERT) by Google introduced a different approach focused on understanding language context by looking both left and right simultaneously. While BERT is primarily designed for comprehension rather than generation, its impact in NLP tasks like question answering and sentiment analysis has been profound.

Variants such as RoBERTa, ALBERT, and DistilBERT optimized BERT’s architecture for better performance, faster training, and smaller sizes.

Other Noteworthy Models

- **T5 (Text-to-Text Transfer Transformer):** Converts every NLP problem into a text-to-text format, making it highly flexible.
- **XLNet:** Combines autoregressive and autoencoding methods, improving on BERT’s limitations.
- **PaLM by Google:** With 540 billion parameters, PaLM pushes the boundaries even further, showing remarkable few-shot learning capabilities.

These models demonstrate the rapid innovation occurring in the field and the variety of approaches researchers are taking to enhance language understanding and generation.

Applications and Impact of Large Language Models

LLMs are not just feats of engineering; they have real-world applications transforming industries and daily life.

Enhancing Communication and Accessibility

From chatbots providing customer support to AI-powered writing assistants improving grammar and style, large language models help bridge communication gaps. They also enable more accessible technologies, such as real-time translation and speech-to-text tools, empowering people across language barriers and disabilities.

Content Creation and Creativity

Writers, marketers, and designers leverage LLMs to brainstorm ideas, draft articles, or generate code snippets. These models act as creative collaborators, accelerating workflows and inspiring new kinds of digital art and storytelling.

Research and Data Analysis

By summarizing large documents, extracting insights, and answering complex queries, LLMs assist researchers in fields ranging from medicine to law. This capability helps sift through vast information quickly, making knowledge more manageable and actionable.

Challenges and Ethical Considerations

While the benefits of large language models are impressive, they also raise ethical questions around bias, misinformation, and privacy. Since LLMs learn from vast datasets scraped from the internet, they can inadvertently reproduce societal biases or generate harmful content.

Developers and organizations must prioritize transparency, fairness, and responsible AI use. Techniques like bias mitigation, content filtering, and user feedback loops are critical to ensuring these models serve society positively.

Future Directions in Large Language Model Research

The future of large language models promises even more exciting developments. Researchers are investigating ways to make models more efficient, interpretable, and aligned with human values.

Multimodal Models

Integrating language with other data types such as images, audio, or video is gaining traction. Multimodal models can understand and generate content that spans multiple senses, opening up new possibilities in virtual assistants, education, and entertainment.

Smaller, Specialized Models

Instead of relying solely on ever-larger general-purpose models, there’s growing interest in creating smaller, domain-specific models that perform better on targeted tasks with less computational overhead.

Improving Explainability

One of the biggest challenges is making model decisions transparent and understandable. Enhanced interpretability will build user trust and allow better debugging and refinement of AI systems.

Human-AI Collaboration

Rather than replacing humans, the focus is shifting toward collaboration—where AI augments human creativity, decision-making, and problem-solving. The evolving relationship between humans and AI promises rich, synergistic outcomes.

Exploring a survey of large language models reveals not only the technical marvels behind these tools but also the profound ways they are changing our interaction with information and each other. As the field advances, staying informed about these developments helps us appreciate both the potential and responsibility that comes with harnessing such powerful AI technologies.

Frequently Asked Questions

What is the primary focus of a survey of large language models?

A survey of large language models primarily focuses on reviewing and analyzing the development, architectures, training methods, capabilities, and applications of large-scale language models in natural language processing.

How have large language models evolved over recent years?

Large language models have evolved from relatively small-scale models to massive transformer-based architectures like GPT, BERT, and their variants, with improvements in scalability, training data size, model parameters, and performance on diverse language tasks.

What are some common challenges highlighted in surveys of large language models?

Common challenges include high computational costs, data biases, ethical concerns, interpretability issues, and the environmental impact of training large models.

What applications of large language models are typically discussed in such surveys?

Surveys usually discuss applications such as machine translation, text generation, question answering, summarization, conversational agents, and code generation.

How do surveys of large language models contribute to future research directions?

These surveys identify current limitations, emerging trends, and potential areas for improvement, guiding researchers in developing more efficient, ethical, and capable language models.

Additional Resources

Surveying the Landscape: A Comprehensive Examination of Large Language Models

a survey of large language models reveals a rapidly evolving field that has become central to advancements in artificial intelligence and natural language processing. These models, characterized by their extensive parameters and vast training datasets, have transformed how machines understand, generate, and interact with human language. From applications in chatbots and content creation to complex problem-solving and language translation, large language models (LLMs) now underpin many AI-driven solutions shaping our digital world.

Understanding Large Language Models

Large language models are sophisticated neural networks trained on enormous corpora of textual data. Their primary objective is to predict the next word or sequence of words in a given context, enabling them to generate coherent and contextually relevant text. Unlike traditional rule-based systems, LLMs learn linguistic patterns, semantics, and syntax implicitly, allowing them to produce human-like language outputs.

The scale of these models is often measured in billions, or even trillions, of parameters. Parameters are the tunable weights within the model that capture linguistic nuances. The growth in parameter count correlates with enhanced model capabilities but also demands significant computational resources. This trade-off between performance and efficiency is a recurring theme in the development and deployment of LLMs.

Key Players in the Large Language Model Arena

Several notable institutions and companies have introduced landmark models that have set benchmarks in natural language understanding and generation:

- **OpenAI’s GPT Series:** The Generative Pre-trained Transformer models, especially GPT-3 and GPT-4, have attracted widespread attention for their versatility and ability to perform few-shot learning, enabling them to adapt to tasks with little explicit training.
- **Google’s PaLM and BERT:** BERT revolutionized language understanding by introducing bidirectional context, improving tasks like question answering and sentiment analysis. PaLM, a more recent model, emphasizes scaling and multitask learning across languages.
- **Meta’s LLaMA:** The Large Language Model Meta AI (LLaMA) series offers competitive performances with fewer parameters, focusing on research accessibility and efficiency.
- **Anthropic’s Claude:** Emphasizing safety and alignment, Claude represents efforts to create more controllable and ethical language models.

Comparative Analysis of Large Language Models

When surveying large language models, it is essential to consider various factors beyond sheer size, including architectural innovations, training methodologies, and application domains.

Model Size vs. Performance

While intuitive to assume that bigger models yield better results, the relationship is nuanced. GPT-3, with 175 billion parameters, demonstrated unprecedented language generation capacity upon release. However, models like Google’s PaLM (540 billion parameters) and GPT-4 deliver enhanced reasoning and contextual understanding, albeit at higher computational costs.

Meta’s LLaMA models challenge the notion that bigger is always better by achieving competitive outcomes with fewer parameters, thanks to optimized training techniques and efficient architecture. This balance is critical for democratizing access to LLM technology in research and industry.

Training Data and Multilingual Capabilities

The scope and diversity of training data significantly impact an LLM’s versatility. Models trained on multilingual datasets, such as PaLM and XLM-R, perform robustly across languages, enabling applications in global contexts. Conversely, models predominantly trained on English corpora may struggle with less-represented languages or dialects.

Moreover, data sources influence model biases and ethical considerations. Curating balanced, high-quality datasets remains a challenge, as LLMs can inadvertently perpetuate stereotypes or misinformation present in their training material.

Applications and Implications of Large Language Models

Large language models have permeated numerous sectors, driving innovation but also raising critical questions about ethics, security, and societal impact.

Diverse Use Cases

- **Content Generation:** From automated article writing to creative storytelling, LLMs assist in producing human-like text efficiently.
- **Conversational Agents:** Chatbots and virtual assistants leverage LLMs to provide more natural and context-aware interactions.
- **Code Synthesis:** Advanced models like OpenAI’s Codex enable automatic code generation, facilitating software development workflows.
- **Language Translation and Summarization:** Multilingual models enhance cross-lingual communication and information digestion.

Challenges and Ethical Considerations

As large language models grow in influence, concerns about their responsible use become paramount. Issues include:

- **Bias and Fairness:** Models may reflect and amplify societal biases embedded in training data.
- **Misinformation Risks:** The ability to generate plausible but false content necessitates robust detection and verification strategies.
- **Environmental Impact:** Training and running massive models consume considerable energy, prompting calls for more sustainable AI practices.
- **Accessibility and Control:** Ensuring equitable access while preventing misuse involves balancing openness with regulatory oversight.

Future Directions in Large Language Model Research

The trajectory of large language models points towards continued scaling, improved alignment, and integration with other modalities such as vision and audio. Techniques like fine-tuning, reinforcement learning from human feedback (RLHF), and model compression aim to enhance usability and reduce deployment costs.

Research is also exploring ways to better interpret and explain model decisions, addressing the “black box” nature of deep learning systems. This transparency is vital for building trust and facilitating collaboration between humans and AI.

In addition, the emergence of open-source LLMs fosters innovation and democratization, allowing a broader community to contribute to and benefit from advances in language technology.

The ongoing survey of large language models underscores their transformative potential and the multifaceted challenges that accompany their rise. As these systems become more integrated into daily life, striking a balance between capability, responsibility, and accessibility will define the next chapter of AI evolution.

A Survey Of Large Language Models

a survey of large language models: Generative AI: Techniques, Models and Applications Rajan Gupta, Sanju Tiwari, Poonam Chaudhary, 2025-03-26 This book unlocks the full potential of modern AI systems through a meticulously structured exploration of concepts, techniques, and practical applications. This comprehensive book bridges theoretical foundations with real-world implementations, offering readers a unique perspective on the rapidly evolving field of generative technologies. From computational foundations to ethical considerations, the book systematically covers essential topics including foundation models, large-scale architectures, prompt engineering, and practical applications. The content seamlessly integrates complex technical concepts with industry-relevant examples, making it an invaluable resource for researchers, academicians, and practitioners. Distinguished by its balanced approach to theory and practice, this book serves as both a learning tool and reference guide. Readers will benefit from: Clear explanations of advanced concepts. Practical implementation insights. Current industry applications. Ethical framework discussions. Whether you're conducting research, implementing solutions, or exploring the field, this book provides the knowledge necessary to understand and apply generative AI technologies effectively while considering crucial aspects of security, privacy, and fairness.

a survey of large language models: *Neutrosophic Sets and Systems, vol. 72/2024* Florentin Smarandache, Mohamed Abdel-Basset, Said Broumi, 2024-11-01 “Neutrosophic Sets and Systems” has been created for publications on advanced studies in neutrosophy, neutrosophic set, neutrosophic logic, neutrosophic probability, neutrosophic statistics that started in 1995 and their applications in any field, such as the neutrosophic structures developed in algebra, geometry, topology, etc. Neutrosophy is a new branch of philosophy that studies the origin, nature, and scope of neutralities, as well as their interactions with different ideational spectra. This theory considers every notion or idea <A> together with its opposite or negation <antiA> and with their spectrum of neutralities <neutA> in between them (i.e. notions or ideas supporting neither <A> nor <antiA>). The <neutA> and <antiA> ideas together are referred to as <nonA>. Neutrosophy is a generalization of Hegel's dialectics (the last one is based on <A> and <antiA> only). According to

this theory every idea <A> tends to be neutralized and balanced by <antiA> and <nonA> ideas - as a state of equilibrium. In a classical way <A>, <neutA>, <antiA> are disjoint two by two. But, since in many cases the borders between notions are vague, imprecise, Sorites, it is possible that <A>, <neutA>, <antiA> (and <nonA> of course) have common parts two by two, or even all three of them as well. Neutrosophic Set and Neutrosophic Logic are generalizations of the fuzzy set and respectively fuzzy logic (especially of intuitionistic fuzzy set and respectively intuitionistic fuzzy logic). In neutrosophic logic a proposition has a degree of truth (T), a degree of indeterminacy (I), and a degree of falsity (F), where T, I, F are standard or non-standard subsets of]-0, 1+[. Neutrosophic Probability is a generalization of the classical probability and imprecise probability. Neutrosophic Statistics is a generalization of the classical statistics.

a survey of large language models: Handbuch Data Science und KI Katherine Munro, Stefan Papp, Zoltan Toth, Wolfgang Weidinger, Danko Nikolic, Barbora Antasova Vesela, Karin Bruckmüller, Annalisa Cadonna, Jana Eder, Jeannette Gorzala, Gerald A. Hahn, Georg Langs, Roxane Licandro, Christian Mata, Sean McIntyre, Mario Meir-Huber, György Móra, Manuel Pasieska, Victoria Rugli, Rania Wazir, Günther Zauner, 2025-02-10 - Bietet einen umfassenden Überblick über die verschiedenen Anwendungsbereiche von Data Science und KI - Mit Fallbeispielen aus der Praxis, um die beschriebenen Konzepte greifbar zu machen - Mit praktischen Beispielen, die Ihnen helfen, einfache Datenanalyseprojekte durchzuführen - Neu in der 3. Auflage: Generativ KI und LLMs, KI und Klimawandel, ML Ops und ML Security, Zahlreiche Kapitel wurden von Grund auf überarbeitet - Ihr exklusiver Vorteil: E-Book inklusive beim Kauf des gedruckten Buches Data Science, Big Data und künstliche Intelligenz gehören derzeit zu den Konzepten, über die in Industrie, Regierung und Gesellschaft viel gesprochen wird, die aber auch am häufigsten missverstanden werden. Dieses Buch erklärt die Konzepte und vermittelt Ihnen das praktische Wissen, um sie zu nutzen. Das Buch nähert sich den Themen Data Science und KI von mehreren Seiten. Es zeigt, wie Sie Data-Plattformen aufbauen und Data-Science-Tools und -Methoden einsetzen können. Auf dem Weg dorthin hilft es Ihnen zu verstehen – und den verschiedenen Interessengruppen zuerklären –, wie Sie mit diesen Techniken Mehrwert generieren können. So kann Data Science in Unternehmen dabei helfen, schnellere Entscheidungen zu treffen, Kosten zu senken und neue Märkte zu erschließen. Darüber hinaus werden die grundlegenden Konzepte von Data Science, einschließlich Statistik, Mathematik sowie rechtliche Überlegungen erklärt. Praktische Fallstudien veranschaulichen, wie aus Daten generiertes Wissen verschiedene Branchen langfristig verändern wird. Das Autor:innenteam besteht aus Datenexpert:innen aus der Wirtschaft und aus dem akademischen Umfeld. Das Spektrum reicht von strategisch ausgerichteten Führungskräften über Data Engineers, die Produktivsysteme erstellen, bis hin zu Data Scientists, die aus Daten Wert generieren. Alle Autor:innen sind im Vorstand oder Mitglieder der Vienna Data Science Group (VDSG). Diese NGO hat sich zum Ziel gesetzt, eine Plattform für den Wissensaustausch zu etablieren. AUS DEM INHALT // - Grundlagen der Mathematik: ML-Algorithmen verstehen und nutzen - Machine Learning: Von statistischen zu neuronalen Verfahren; von Transformers und GPT-3 bis AutoML - Natural Language Processing: Werkzeuge und Techniken zur Gewinnung von Erkenntnissen aus Textdaten und zur Entwicklung von Sprachtechnologien - Computer Vision: Erkenntnisse aus Bildern und Videos gewinnen - Modellierung und Simulation: Modellierung des Verhaltens komplexer Systeme, z. B. der Ausbreitung von COVID-19. Was-wäre-wenn-Analysen - ML und KI in der Produktion: Vom Experiment zum Data- Science-Produkt - Ergebnisse präsentieren: Grundlegende Präsentationstechniken für Data Scientists

a survey of large language models: An Efficient SuperHyperSoft Framework for Evaluating LLMs-based Secure Blockchain Platforms Mona Mohamed, Alaa Elmor, Florentin Smarandache, Ahmed A. Metwaly, 2024-01-01 In the age of modern technology, the use of the internet has become imperative. However, this widespread access presents a double-edged sword and opens doors for hackers and scammers to exploit vulnerabilities and engage in illegal activities. Accordingly, scholars and stakeholders are attempting to solve this matter. Large Language Models (LLMs) have provided highly effective methodologies and solutions in various cybersecurity sectors.

Hence, we exhibited the efficacy of LLMs in several information and communication technologies (ICT) such as the Internet of Things (IoT), cloud computing, blockchain technology (BCT), etc. which are attacked and threatened. Accordingly, the objective of our study is to illustrate how LLMs are supporting ICT, especially BC to be secure against attacks. Another study's objective is to aid the stakeholders and enterprises that seek resilience and sustainability by recommending the most secure BC platform to adopt in critical sectors. Wherein, LLMs support BC in many directions as developing secure smart contracts and scanning the smart contract to protect it from any subversive acts by identifying anomalous activities. Hence, we suggested a soft opting model to rank the alternatives of BC platforms and recommend optimal BC. Also, the process of constructing this model requires leveraging several techniques. We applied for the first time SuperHyperSoft (SHS) as an extension of Hypersoft to treat various attributes and sub-attributes for BC based on LLMs. Multi-criteria decision-making (MCDM)techniques are utilized for their ability to treat conflicting sub-attributes. Hence, entropy and multi-objective optimization based on simple ratio analysis (MOOSRA) are utilized as techniques of MCDM. These techniques are working under the authority of the Single Value Neutrosophic (SVN) technique to support MCDM techniques in ambiguous situations.

a survey of large language models: *Finance and Large Language Models* Paul Moon Sub Choi, Seth H. Huang, 2025-08-21 This book highlights how AI agents and Large Language Models (LLMs) are set to revolutionize the finance and trading sectors in unprecedented ways. These technologies bring a new level of sophistication to data analysis and decision-making, enabling real-time processing of vast and complex datasets with unparalleled accuracy and speed. AI agents, equipped with advanced machine learning algorithms, can identify patterns and predict market trends with a level of precision that may soon surpass human capabilities. LLMs, on the other hand, facilitate the interpretation and synthesis of unstructured data, such as financial news, reports, and social media sentiments, providing deeper insights and more informed trading strategies. This convergence of AI and LLM technology not only enhances the efficiency and profitability of trading operations but also introduces a paradigm shift in risk management, compliance, and personalized financial services. As these technologies continue to evolve, they promise to democratize access to sophisticated trading tools and insights, leveling the playing field for individual traders and smaller financial institutions while driving innovation and growth across the entire financial ecosystem.

a survey of large language models: HCI International 2023 – Late Breaking Posters Constantine Stephanidis, Margherita Antona, Stavroula Ntoa, Gavriel Salvendy, 2023-12-11 This two-volme set CCIS 1957-1958 is part of the refereed proceedings of the 25th International Conference on Human-Computer Interaction, HCII 2023, which was held in Copenhagen, Denmark, in July 2023. A total of 5583 individuals from academia, research institutes, industry, and governmental agencies from 88 countries submitted contributions, and 1276 papers and 275 posters were included in the proceedings that were published just before the start of the conference. Additionally, 296 papers and 181 posters are included in the volumes of the proceedings published after the conference, as "Late Breaking Work" (papers and posters). The contributions thoroughly cover the entire field of human-computer interaction, addressing major advances in knowledge and effective use of computers in a variety of application areas.

a survey of large language models: *Foundations and Practice of Security* Kamel Adi, Simon Bourdeau, Christel Durand, Valérie Viet Triem Tong, Alina Dulipovici, Yvon Kermarrec, Joaquin Garcia-Alfaro, 2025-04-30 This two-volume set constitutes the refereed proceedings of the 17th International Symposium on Foundations and Practice of Security, FPS 2024, held in Montréal, QC, Canada, during December 09–11, 2024. The 28 full and 11 short papers presented in this book were carefully reviewed and selected from 75 submissions. The papers were organized in the following topical sections: Part I: Critical issues of protecting systems against digital threats,considering financial, technological, and operational implications; Automating and enhancing security mechanisms in software systems and data management; Cybersecurity and AI when applied to emerging technologies; Cybersecurity and Ethics; Cybersecurity and privacy in connected and

autonomous systems for IoT, smart environments, and critical infrastructure; New trends in advanced cryptographic protocols. Part II: Preserving privacy and maintaining trust for end users in a complex and numeric cyberspace; Intersecting security, privacy, and machine learning techniques to detect, mitigate, and prevent threats; New trends of machine leaning and AI applied to cybersecurity.

a survey of large language models: Intelligence Science V Zhongzhi Shi, Michael Witbrock, Qi Tian, 2024-09-28 This book constitutes the refereed proceedings of the 6th IFIP TC 12 International Conference on Intelligence Science, ICIS 2024, held in Nanjing, China, in October 25-28, 2024. The 23 full papers and 2 short papers presented here were carefully reviewed and selected from 32 submissions. These papers have been categorized into the following sections: Machine Learning; Causal Reasoning; Large Language Model; Intelligent Robot; Perceptual Intelligence; AI for Science; Medical Artificial Intelligence.

a survey of large language models: Advanced Intelligent Computing Technology and Applications De-Shuang Huang, Haiming Chen, Bo Li, Qinhu Zhang, 2025-07-21 The 12-volume set CCIS 2564-2575, together with the 28-volume set LNCS/LNAI/LNBI 15842-15869, constitutes the refereed proceedings of the 21st International Conference on Intelligent Computing, ICIC 2025, held in Ningbo, China, during July 26-29, 2025. The 523 papers presented in these proceedings books were carefully reviewed and selected from 4032 submissions. This year, the conference concentrated mainly on the theories and methodologies as well as the emerging applications of intelligent computing. Its aim was to unify the picture of contemporary intelligent computing techniques as an integral concept that highlights the trends in advanced computational intelligence and bridges theoretical research with applications. Therefore, the theme for this conference was Advanced Intelligent Computing Technology and Applications.

a survey of large language models: *Computer Vision – ECCV 2024* Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, Gül Varol, 2024-11-26 The multi-volume set of LNCS books with volume numbers 15059 up to 15147 constitutes the refereed proceedings of the 18th European Conference on Computer Vision, ECCV 2024, held in Milan, Italy, during September 29–October 4, 2024. The 2387 papers presented in these proceedings were carefully reviewed and selected from a total of 8585 submissions. The papers deal with topics such as computer vision; machine learning; deep neural networks; reinforcement learning; object recognition; image classification; image processing; object detection; semantic segmentation; human pose estimation; 3d reconstruction; stereo vision; computational photography; neural networks; image coding; image reconstruction; motion estimation.

a survey of large language models: *Challenges in Large Language Model Development and AI Ethics* Gupta, Brij, 2024-08-15 The development of large language models has resulted in artificial intelligence advancements promising transformations and benefits across various industries and sectors. However, this progress is not without its challenges. The scale and complexity of these models pose significant technical hurdles, including issues related to bias, transparency, and data privacy. As these models integrate into decision-making processes, ethical concerns about their societal impact, such as potential job displacement or harmful stereotype reinforcement, become more urgent. Addressing these challenges requires a collaborative effort from business owners, computer engineers, policymakers, and sociologists. Fostering effective research for solutions to address AI ethical challenges may ensure that large language model developments benefit society in a positive way. Challenges in Large Language Model Development and AI Ethics addresses complex ethical dilemmas and challenges of the development of large language models and artificial intelligence. It analyzes ethical considerations involved in the design and implementation of large language models, while exploring aspects like bias, accountability, privacy, and social impacts. This book covers topics such as law and policy, model architecture, and machine learning, and is a useful resource for computer engineers, sociologists, policymakers, business owners, academicians, researchers, and scientists.

a survey of large language models: Biocomputing 2025 - Proceedings Of The Pacific

Symposium Russ B Altman, Lawrence Hunter, Marylyn D Ritchie, Tiffany A Murray, Teri E Klein, 2024-11-29 The Pacific Symposium on Biocomputing (PSB) 2025 is an international, multidisciplinary conference for the presentation and discussion of current research in the theory and application of computational methods in problems of biological significance. Presentations are rigorously peer reviewed and are published in an archival proceedings volume. PSB 2025 will be held on January 4 - 8, 2025 in Kohala Coast, Hawaii. Tutorials and workshops will be offered prior to the start of the conference.PSB 2025 will bring together top researchers from the US, the Asian Pacific nations, and around the world to exchange research results and address open issues in all aspects of computational biology. It is a forum for the presentation of work in databases, algorithms, interfaces, visualization, modeling, and other computational methods, as applied to biological problems, with emphasis on applications in data-rich areas of molecular biology.The PSB has been designed to be responsive to the need for critical mass in sub-disciplines within biocomputing. For that reason, it is the only meeting whose sessions are defined dynamically each year in response to specific proposals. PSB sessions are organized by leaders of research in biocomputing's 'hot topics.' In this way, the meeting provides an early forum for serious examination of emerging methods and approaches in this rapidly changing field.

a survey of large language models: *Applied Machine Learning and Deep Learning: Architectures and Techniques* Nitin Liladhar Rane, Suraj Kumar Mallick, Ömer Kaya, Jayesh Rane, 2024-10-13 This book provides an extensive overview of recent advances in machine learning (ML) and deep learning (DL). It starts with a comprehensive introduction to the latest architectural and design practices, with an overview of basic techniques and optimization algorithms and methodologies that are fundamental to modern ML/DL development followed by the tools and frameworks that are driving innovation in ML/DL. The presentation then points to the central position of ML and DL in developing generative AI like ChatGPT. Then look at different industrial applications such as explaining the real-world impacts of each. This includes challenges around corroborate artificial Intelligence (AI), and trustworthy AI, and so on. Finally, the book presents a futuristic vision on the potentials and implications of future ML and DL architectures, making it an ideal guide for researchers, practitioners and industry professionals. This book will be a significant resource for comprehending present advancements, addressing encounter challenges, and traversing the ML and DL landscape in future, making it an indispensable reference for anyone interested in applying these technologies across sectors.

a survey of large language models: Hybrid Artificial Intelligent Systems Héctor Quintián, Emilio Corchado, Alicia Troncoso Lora, Hilde Pérez García, Esteban Jove Pérez, José Luis Calvo Rolle, Francisco Javier Martínez de Pisón, Pablo García Bringas, Francisco Martínez Álvarez, Álvaro Herrero, Paolo Fosci, 2024-10-08 This two-part proceedings volume constitutes the refereed proceedings of the 19th International Conference on Hybrid Artificial Intelligent Systems (HAIS 2024), held in Salamanca, Spain, during October 9–11, 2024. The 52 full papers were carefully reviewed and selected from 112 submissions. They were organized in topical sections as follows: Part I: Biomedical Applications, Data Mining and Decision Support Systems, Deep Learning, Evolutionary Computation and Optimization. Part II: HAIS Applications, HAIS Energy Applications, Image and Text Processing, Reinforced Learning.

a survey of large language models: Recent Advances in Machine Learning Techniques and Sensor Applications for Human Emotion, Activity Recognition and Support Kyandoghere Kyamakya, Fadi Al Machot, Habib Ullah, Florenc Demrozi, 2024-11-07 This book explores integrating machine learning techniques and sensor applications for human emotion and activity recognition, creating personalized and effective support systems. It covers state-of-the-art machine learning techniques and large language models using multimodal sensors. Enhancing the quality of life for individuals with special needs, particularly the elderly, is a key focus in Active and Assisted Living (AAL) research. Unlike other literature, it emphasizes support mechanisms along with recognition, using metamodel integration for adaptable AAL systems. This book offers insights into technologies transforming AAL for researchers, students, and practitioners. It is a valuable resource

for developing responsive and personalized support systems that enhance life quality in smart environments. It is also essential for advancing the understanding of machine learning and sensor technologies in AAL and emotion recognition.

a survey of large language models: Information Access in the Era of Generative AI Ryen W. White, Chirag Shah, 2024-12-24 Generative Artificial Intelligence (GenAI) has emerged as a groundbreaking technology that promises to revolutionize many industries as well as people's personal and professional lives. This book discusses GenAI and its role in information access - often referred to as Generative Information Retrieval (GenIR) - or more broadly, information interaction. The role of GenAI in information access is complex and dynamic, with many dimensions. To address this, following a brief introduction to GenAI and GenIR, the remainder of the book provides eight chapters, each targeting a different dimension or sub-topic. These cover foundations of GenIR, interactions with GenIR systems, adapting them to users, tasks, and scenarios, improving them based on user feedback, GenIR evaluation, the sociotechnical implications of GenAI for information access, recommendations within GenIR, and the future of information access with GenIR. The book is targeted at graduate students and researchers interested in issues of information retrieval, access, and interactions, as well as applications of GenAI in various informational contexts. While some of the parts assume prior background in IR or AI, most others do not, making this book suitable for adoption in various classes as a primary source or as a supplementary material.

a survey of large language models: KI in der Projektwirtschaft 2 Christian Bernert, Steffen Scheurer, Harald Wehnes, 2024-09-23 Dieser Band zeigt, wie KI die Projektwirtschaft neugestaltet werden kann und welche Herausforderungen in Projekten sich damit innovativ und effizient meistern lassen. Mit den Schwerpunktthemen KI-Einsatz in konkreten Anwendungen, in der Forschung zu KI Start-ups für Projektmanagement, Veränderung der Rollen in den Unternehmen durch KI sowie Portfolio und Mustererkennung wird der Fokus auf konkrete praxisnahe Anwendung gelegt. Es wird dargelegt, wie erfolgreiche Umsetzungen aus der Forschung in die Anwendung durch Start-ups gelingen können. Darüber hinaus entstehen mit der Einführung von KI in die Projektwirtschaft neue Berufsbilder und Rollen, wie z.B. Datenanalyst:innen für Projekte, KI-Strateg:innen oder spezialisierte KI-Trainer:innen, die eine Neuausrichtung der Aus- und Weiterbildung und Anpassungen in der Unternehmenskultur erfordern. Als innovativer Anwendungsbereich von KI in der Projektwirtschaft wird der Bereich Early Warning Systems gesehen, indem mittels Mustererkennung gezieltes präventives Handeln ermöglicht und damit bessere Projektergebnisse erzielt werden können.

a survey of large language models: Testing Software and Systems Héctor D. Menéndez, Gema Bello-Orgaz, Pepita Barnard, John Robert Bautista, Arya Farahi, Santanu Dash, DongGyun Han, Sophie Fortz, Victor Rodriguez-Fernandez, 2025-01-24 This book constitutes the refereed proceedings of the 36th IFIP WG 6.1 International Conference on Testing Software and Systems, ICTSS 2024, held in London, UK, during October 30–November 1, 2024. The 17 full papers and 5 short papers included in this book were carefully reviewed and selected from 40 submissions. They were organized in topical sections as follows: Best Paper Award; Industry and Challenge Tracks; Mutation Testing and Code Generation; Advancing Code Vulnerability Detection; Short Papers; Tutorial; Journal First; Health Track; Innovations in Software Testing and AI Compliance; Improving Software Testing Reliability and Advancements in Testing Methodologies.

a survey of large language models: Building AI Agents with LLMs, RAG, and Knowledge Graphs Salvatore Raieli, Gabriele Iuculano, 2025-07-11 Master LLM fundamentals to advanced techniques like RAG, reinforcement learning, and knowledge graphs to build, deploy, and scale intelligent AI agents that reason, retrieve, and act autonomously Key Features Implement RAG and knowledge graphs for advanced problem-solving Leverage innovative approaches like LangChain to create real-world intelligent systems Integrate large language models, graph databases, and tool use for next-gen AI solutions Purchase of the print or Kindle book includes a free PDF eBook Book DescriptionThis AI agents book addresses the challenge of building AI that not only generates text but also grounds its responses in real data and takes action. Authored by AI specialists with deep

expertise in drug discovery and systems optimization, this guide empowers you to leverage retrieval-augmented generation (RAG), knowledge graphs, and agent-based architectures to engineer truly intelligent behavior. By combining large language models (LLMs) with up-to-date information retrieval and structured knowledge, you'll create AI agents capable of deeper reasoning and more reliable problem-solving. Inside, you'll find a practical roadmap from concept to implementation. You'll discover how to connect language models with external data via RAG pipelines for increasing factual accuracy and incorporate knowledge graphs for context-rich reasoning. The chapters will help you build and orchestrate autonomous agents that combine planning, tool use, and knowledge retrieval to achieve complex goals. Concrete Python examples built on popular libraries, along with real-world case studies, reinforce each concept and show you how these techniques come together. By the end of this book, you'll be well-equipped to build intelligent AI agents that reason, retrieve, and interact dynamically, empowering you to deploy powerful AI solutions across industries.What you will learn Learn how LLMs work, their structure, uses, and limits, and design RAG pipelines to link them to external data Build and query knowledge graphs for structured context and factual grounding Develop AI agents that plan, reason, and use tools to complete tasks Integrate LLMs with external APIs and databases to incorporate live data Apply techniques to minimize hallucinations and ensure accurate outputs Orchestrate multiple agents to solve complex, multi-step problems Optimize prompts, memory, and context handling for long-running tasks Deploy and monitor AI agents in production environments Who this book is for If you are a data scientist or researcher who wants to learn how to create and deploy an AI agent to solve limitless tasks, this book is for you. To get the most out of this book, you should have basic knowledge of Python and Gen AI. This book is also excellent for experienced data scientists who want to explore state-of-the-art developments in LLM and LLM-based applications.

a survey of large language models: Multimodal Generative AI Akansha Singh, Krishna Kant Singh, 2025-02-24 This book stands at the forefront of AI research, offering a comprehensive examination of multimodal generative technologies. Readers are taken on a journey through the evolution of generative models, from early neural networks to contemporary marvels like GANs and VAEs, and their transformative application in synthesizing realistic images and videos. In parallel, the text delves into the intricacies of language models, with a particular on revolutionary transformer-based designs. A core highlight of this work is its detailed discourse on integrating visual and textual models, laying out state-of-the-art techniques for creating cohesive, multimodal AI systems. "Multimodal Generative AI" is more than a mere academic text; it's a visionary piece that speculates on the future of AI, weaving through case studies in autonomous systems, content creation, and human-computer interaction. The book also fosters a dialogue on responsible innovation in this dynamic field. Tailored for postgraduates, researchers, and professionals, this book is a must-read for anyone vested in the future of AI. It empowers its readers with the knowledge to harness the potential of multimodal systems in solving complex problems, merging visual understanding with linguistic prowess. This book can be used as a reference for postgraduates and researchers in related areas.

Related to a survey of large language models

Create a survey - Google Surveys Help Where will my survey questions appear? Questions appear throughout sites in our publisher network in order to get a representative sample of respondents. Users complete survey

Create your first form in Google Forms When someone takes your survey, they will be required to enter their email address before they submit the form. Collect verified emails Important: Respondents must confirm their Google

Quick Start Guide - Google Surveys Help How to set up screening questions Select the checkbox for each answer that qualifies a respondent for this audience. Having three or more answers helps eliminate

Crear una encuesta - Ayuda de Surveys - Google Help Cuando Google Surveys recoge

respuestas de la "audiencia general de Internet", utiliza conjuntos de datos de población de Internet publicados para realizar la distribución de la

Google Surveys Sunset Historical survey results downloads are no longer available. We began Surveys over 10 years ago to enable businesses of all sizes to run custom market research with an

Device Usage Study Help - Google Help Official Device Usage Study Help Help Center where you can find tips and tutorials on using Device Usage Study Help and other answers to frequently asked questions

Google Surveys Help Official Google Surveys Help Center where you can find tips and tutorials on using Google Surveys and other answers to frequently asked questions

View and export results - Google Surveys Help To view your survey results: Sign in to Google Surveys. Click the survey you want to view on the survey dashboard. Click the text of any question to see individual question results. Keep in

Earn rewards - Opinion Rewards Help - Google Help With Google Opinion Rewards, you'll take surveys that are run by market researchers. Survey frequency may vary, and you don't have to answer every survey you receive. In exchange,

Earn rewards - Google Surveys Help The amount of Google Play credits you receive for answering surveys can vary and is dependent on the number of questions in the survey and the amount of time it takes to answer the survey.

Create a survey - Google Surveys Help Where will my survey questions appear? Questions appear throughout sites in our publisher network in order to get a representative sample of respondents. Users complete survey

Create your first form in Google Forms When someone takes your survey, they will be required to enter their email address before they submit the form. Collect verified emails Important: Respondents must confirm their Google

Quick Start Guide - Google Surveys Help How to set up screening questions Select the checkbox for each answer that qualifies a respondent for this audience. Having three or more answers helps eliminate

Crear una encuesta - Ayuda de Surveys - Google Help Cuando Google Surveys recoge respuestas de la "audiencia general de Internet", utiliza conjuntos de datos de población de Internet publicados para realizar la distribución de la

Google Surveys Sunset Historical survey results downloads are no longer available. We began Surveys over 10 years ago to enable businesses of all sizes to run custom market research with an

Device Usage Study Help - Google Help Official Device Usage Study Help Help Center where you can find tips and tutorials on using Device Usage Study Help and other answers to frequently asked questions

Google Surveys Help Official Google Surveys Help Center where you can find tips and tutorials on using Google Surveys and other answers to frequently asked questions

View and export results - Google Surveys Help To view your survey results: Sign in to Google Surveys. Click the survey you want to view on the survey dashboard. Click the text of any question to see individual question results. Keep in

Earn rewards - Opinion Rewards Help - Google Help With Google Opinion Rewards, you'll take surveys that are run by market researchers. Survey frequency may vary, and you don't have to answer every survey you receive. In exchange,

Earn rewards - Google Surveys Help The amount of Google Play credits you receive for answering surveys can vary and is dependent on the number of questions in the survey and the amount of time it takes to answer the survey.

Create a survey - Google Surveys Help Where will my survey questions appear? Questions appear throughout sites in our publisher network in order to get a representative sample of respondents. Users complete survey

Create your first form in Google Forms When someone takes your survey, they will be required to enter their email address before they submit the form. Collect verified emails Important:

Respondents must confirm their Google

Quick Start Guide - Google Surveys Help How to set up screening questions Select the checkbox for each answer that qualifies a respondent for this audience. Having three or more answers helps eliminate

Crear una encuesta - Ayuda de Surveys - Google Help Cuando Google Surveys recoge respuestas de la "audiencia general de Internet", utiliza conjuntos de datos de población de Internet publicados para realizar la distribución de la

Google Surveys Sunset Historical survey results downloads are no longer available. We began Surveys over 10 years ago to enable businesses of all sizes to run custom market research with an

Device Usage Study Help - Google Help Official Device Usage Study Help Help Center where you can find tips and tutorials on using Device Usage Study Help and other answers to frequently asked questions

Google Surveys Help Official Google Surveys Help Center where you can find tips and tutorials on using Google Surveys and other answers to frequently asked questions

View and export results - Google Surveys Help To view your survey results: Sign in to Google Surveys. Click the survey you want to view on the survey dashboard. Click the text of any question to see individual question results. Keep in

Earn rewards - Opinion Rewards Help - Google Help With Google Opinion Rewards, you'll take surveys that are run by market researchers. Survey frequency may vary, and you don't have to answer every survey you receive. In exchange,

Earn rewards - Google Surveys Help The amount of Google Play credits you receive for answering surveys can vary and is dependent on the number of questions in the survey and the amount of time it takes to answer the survey.

Related to a survey of large language models

What Are Large Language Models? Definition, Examples & Future Of LLMS (The Next Hint12d) What are LLMs? Know their working, meaning, benefits, & application, and discover the best large language model examples

What Are Large Language Models? Definition, Examples & Future Of LLMS (The Next Hint12d) What are LLMs? Know their working, meaning, benefits, & application, and discover the best large language model examples

How to Make Large Models Smarter? Scientists Reveal Key Deficiencies in Logical Reasoning of Large Models (21d) In recent years, despite large models demonstrating excellent performance in natural language processing tasks, a significant issue that cannot be overlooked is their considerable deficiencies in

How to Make Large Models Smarter? Scientists Reveal Key Deficiencies in Logical Reasoning of Large Models (21d) In recent years, despite large models demonstrating excellent performance in natural language processing tasks, a significant issue that cannot be overlooked is their considerable deficiencies in

UNC's lack of a standardized AI policy leaves students and faculty to their own devices (The Daily Tar Heel1d) In a 2024 survey from the Provost’s AI Committee, 94 percent of undergraduate respondents said that they understood the

UNC's lack of a standardized AI policy leaves students and faculty to their own devices (The Daily Tar Heel1d) In a 2024 survey from the Provost’s AI Committee, 94 percent of undergraduate respondents said that they understood the

Open vs. Closed: The Battle for the Future of Language Models (Aclu.org21d) I’ve been writing about the democratic future of large language models (LLMs). Will this tech turn out to be an inherently centralized, authoritarian technology like nuclear power, or a more

Open vs. Closed: The Battle for the Future of Language Models (Aclu.org21d) I’ve been writing about the democratic future of large language models (LLMs). Will this tech turn out to be an inherently centralized, authoritarian technology like nuclear power, or a more

For DOD, the future of large language models is smaller (Defense One4mon) The U.S. military is working on ways to get the power of cloud-based, big-data AI in tools that can run on local computers, draw upon more focused data sets, and remain safe from spying eyes,
For DOD, the future of large language models is smaller (Defense One4mon) The U.S. military is working on ways to get the power of cloud-based, big-data AI in tools that can run on local computers, draw upon more focused data sets, and remain safe from spying eyes,
Beyond Autoregression: A New Model For Text Generation (7d) There are also trade-offs in creativity. Because the energy critic favors low-energy (i.e., high-probability) text, the model
Beyond Autoregression: A New Model For Text Generation (7d) There are also trade-offs in creativity. Because the energy critic favors low-energy (i.e., high-probability) text, the model
AI models pass CFA Level III exam in minutes, raising questions for advisors (InvestmentNews4d) Research reveals a quantum leap in AI's reasoning ability, renewing the debate around the value of human advisors
AI models pass CFA Level III exam in minutes, raising questions for advisors (InvestmentNews4d) Research reveals a quantum leap in AI's reasoning ability, renewing the debate around the value of human advisors
Do Large Language Models Dream of AI Agents? (1mon) For AI models, knowing what to remember might be as important as knowing what to forget. Welcome to the era of “sleeptime compute.”
Do Large Language Models Dream of AI Agents? (1mon) For AI models, knowing what to remember might be as important as knowing what to forget. Welcome to the era of “sleeptime compute.”

Related Articles

- scott foresman math grade 4
- data nugget urbanization and estuary eutrophication answer key
- lamar odom dating history

Back to Home