Extracting Training Data From Large Language Models

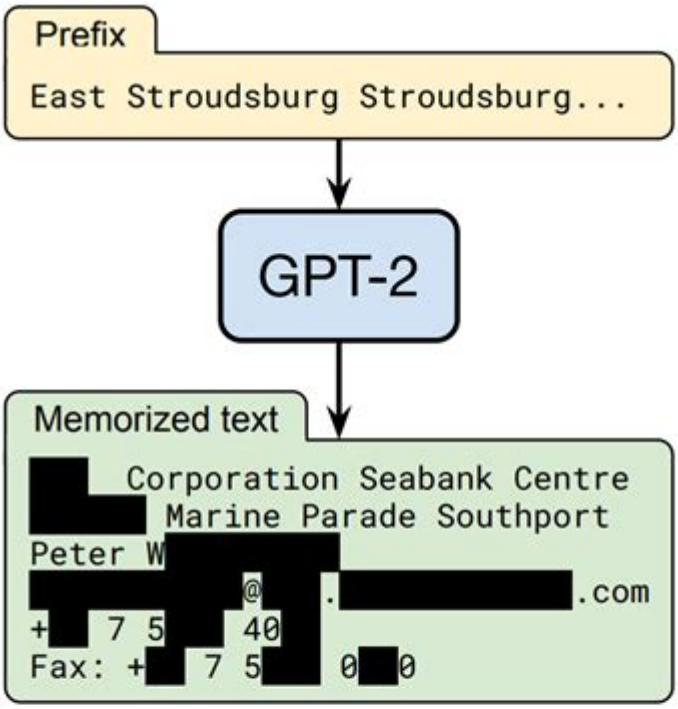

Figure 1: **Our extraction attack.** Given query access to a neural network language model, we extract an individual person's name, email address, phone number, fax number, and physical address. The example in this figure shows information that is all accurate so we redact it to protect privacy.

Extracting Training Data from Large Language Models: A Deep Dive

Large language models (LLMs) have revolutionized the field of artificial intelligence, powering everything from sophisticated chatbots to advanced text generation tools. But have you ever wondered about the secrets behind their impressive capabilities? A crucial aspect is the vast amount of training data used to build them. This post will delve into the complex and often challenging process of extracting training data from large language models, exploring the techniques, challenges, and ethical considerations involved. We'll equip you with a comprehensive understanding of this emerging field, offering insights into both practical methods and future research directions.

Why Extract Training Data from LLMs?

Understanding the data used to train an LLM is crucial for several reasons:

Bias Detection and Mitigation: LLMs can inherit biases present in their training data, leading to unfair or discriminatory outputs. Analyzing the training data allows researchers to identify and mitigate these biases, improving the fairness and equity of the model.

Model Interpretability and Explainability: Knowing the data sources provides insights into how the model makes predictions, enhancing transparency and facilitating trust. This is particularly important in high-stakes applications like healthcare and finance.

Improving Model Performance: Analyzing the training data can reveal weaknesses and gaps, allowing developers to improve the model's performance by augmenting the data or refining the training process.

Copyright and Intellectual Property: Understanding the sources of the training data is essential for addressing potential copyright infringement and intellectual property concerns.

Data Provenance and Auditing: Tracking the origin and usage of training data is crucial for responsible AI development, allowing for better accountability and auditability.

Methods for Extracting Training Data from LLMs

Directly extracting the raw training data used to train a proprietary LLM is often impossible. The data is usually considered confidential and protected by intellectual property rights. However, several indirect methods can provide valuable insights:

1. Analyzing Model Outputs:

By carefully crafting prompts and analyzing the model's responses, we can infer aspects of its training data. For example, observing the model's stylistic choices, factual knowledge, and common themes can reveal clues about the types of text it was trained on. This method is limited but can provide valuable qualitative insights.

2. Membership Inference Attacks:

These attacks attempt to determine whether a specific data point was part of the training dataset. While sophisticated, these methods are computationally intensive and face limitations in their accuracy.

3. Data Leakage Detection:

This involves analyzing the model's outputs for unintentional leakage of information from the training data. This could manifest as verbatim reproduction of sentences or specific phrases from

the training set.

4. Proxy Data Analysis:

If the LLM was trained on publicly available datasets, analyzing these datasets can provide indirect insights into the model's training process. This involves comparing the model's capabilities with the characteristics of the publicly available data.

5. Model Inversion:

This technique aims to reconstruct aspects of the training data by observing the model's behavior. It’s a complex process that requires advanced techniques and often provides only partial reconstructions of the training data.

Challenges and Ethical Considerations

Extracting training data from LLMs presents numerous challenges:

Computational Cost: Many techniques are computationally expensive, requiring significant resources and expertise.

Data Privacy: Accessing and analyzing training data raises significant privacy concerns, especially if personally identifiable information is involved.

Intellectual Property Rights: Using or sharing extracted data may infringe on copyright or other intellectual property rights.

Interpretability Limitations: The relationship between the training data and the model's behavior is complex and not always easily interpretable.

The Future of Extracting Training Data from LLMs

Research into extracting training data from LLMs is an active and evolving field. Future work will likely focus on developing more sophisticated and efficient techniques, addressing ethical concerns, and improving the interpretability of the results. This research is vital for ensuring the responsible and ethical development and deployment of LLMs.

Conclusion

Extracting training data from large language models presents both significant opportunities and challenges. Understanding the methods, limitations, and ethical implications is crucial for researchers, developers, and policymakers alike. As LLMs become increasingly powerful and pervasive, the ability to analyze their training data will become increasingly important for ensuring their responsible and beneficial use.

FAQs

1. Can I directly access the training data of a commercially available LLM? No, the training data of commercially available LLMs is generally proprietary and not publicly accessible.

2. Are membership inference attacks always successful? No, the success of membership inference attacks depends on various factors, including the model's architecture, the size of the training data, and the sophistication of the attack.

3. What are the legal implications of extracting training data? The legal implications depend on various factors, including the source of the data, the methods used for extraction, and the intended use of the extracted data. It is crucial to consult legal counsel before undertaking any data extraction efforts.

4. How can I contribute to research in this area? You can contribute by exploring novel techniques for data extraction, developing tools for analyzing training data, or conducting ethical assessments of LLM training practices.

5. What are the key ethical considerations when working with LLM training data? Key ethical considerations include data privacy, intellectual property rights, bias mitigation, and transparency. A responsible approach requires careful consideration of these factors at every stage of the research process.

extracting training data from large language models: *Machine Learning and Knowledge Extraction* Andreas Holzinger, Peter Kieseberg, Federico Cabitza, Andrea Campagner, A Min Tjoa, Edgar Weippl, 2023-08-21 This volume LNCS-IFIP constitutes the refereed proceedings of the 7th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2023 in Benevento, Italy, during August 28 – September 1, 2023. The 18 full papers presented together were carefully reviewed and selected from 30 submissions. The conference focuses on integrative machine learning approach, considering the importance of data science and visualization for the algorithmic pipeline with a strong emphasis on privacy, data protection, safety and security.

extracting training data from large language models: *Large Language Models* Uday Kamath, Kevin Keenan, Garrett Somers, Sarah Sorenson, 2024 Large Language Models (LLMs) have emerged as a cornerstone technology, transforming how we interact with information and redefining the boundaries of artificial intelligence. LLMs offer an unprecedented ability to understand, generate, and interact with human language in an intuitive and insightful manner, leading to transformative applications across domains like content creation, chatbots, search engines, and research tools. While fascinating, the complex workings of LLMs -- their intricate architecture, underlying algorithms, and ethical considerations -- require thorough exploration, creating a need for a comprehensive book on this subject. This book provides an authoritative exploration of the

design, training, evolution, and application of LLMs. It begins with an overview of pre-trained language models and Transformer architectures, laying the groundwork for understanding prompt-based learning techniques. Next, it dives into methods for fine-tuning LLMs, integrating reinforcement learning for value alignment, and the convergence of LLMs with computer vision, robotics, and speech processing. The book strongly emphasizes practical applications, detailing real-world use cases such as conversational chatbots, retrieval-augmented generation (RAG), and code generation. These examples are carefully chosen to illustrate the diverse and impactful ways LLMs are being applied in various industries and scenarios. Readers will gain insights into operationalizing and deploying LLMs, from implementing modern tools and libraries to addressing challenges like bias and ethical implications. The book also introduces the cutting-edge realm of multimodal LLMs that can process audio, images, video, and robotic inputs. With hands-on tutorials for applying LLMs to natural language tasks, this thorough guide equips readers with both theoretical knowledge and practical skills for leveraging the full potential of large language models. This comprehensive resource is appropriate for a wide audience: students, researchers and academics in AI or NLP, practicing data scientists, and anyone looking to grasp the essence and intricacies of LLMs.

extracting training data from large language models: *Large Language Models in Cybersecurity* Andrei Kucharavy, 2024 This open access book provides cybersecurity practitioners with the knowledge needed to understand the risks of the increased availability of powerful large language models (LLMs) and how they can be mitigated. It attempts to outrun the malicious attackers by anticipating what they could do. It also alerts LLM developers to understand their work's risks for cybersecurity and provides them with tools to mitigate those risks. The book starts in Part I with a general introduction to LLMs and their main application areas. Part II collects a description of the most salient threats LLMs represent in cybersecurity, be they as tools for cybercriminals or as novel attack surfaces if integrated into existing software. Part III focuses on attempting to forecast the exposure and the development of technologies and science underpinning LLMs, as well as macro levers available to regulators to further cybersecurity in the age of LLMs. Eventually, in Part IV, mitigation techniques that should allowsafe and secure development and deployment of LLMs are presented. The book concludes with two final chapters in Part V, one speculating what a secure design and integration of LLMs from first principles would look like and the other presenting a summary of the duality of LLMs in cyber-security. This book represents the second in a series published by the Technology Monitoring (TM) team of the Cyber-Defence Campus. The first book entitled Trends in Data Protection and Encryption Technologies appeared in 2023. This book series provides technology and trend anticipation for government, industry, and academic decision-makers as well as technical experts.

extracting training data from large language models: Adversarial AI Attacks, Mitigations, and Defense Strategies John Sotiropoulos, 2024-07-26 Understand how adversarial attacks work against predictive and generative AI, and learn how to safeguard AI and LLM projects with practical examples leveraging OWASP, MITRE, and NIST Key Features Understand the connection between AI and security by learning about adversarial AI attacks Discover the latest security challenges in adversarial AI by examining GenAI, deepfakes, and LLMs Implement secure-by-design methods and threat modeling, using standards and MLSecOps to safeguard AI systems Purchase of the print or Kindle book includes a free PDF eBook Book DescriptionAdversarial attacks trick AI systems with malicious data, creating new security risks by exploiting how AI learns. This challenges cybersecurity as it forces us to defend against a whole new kind of threat. This book demystifies adversarial attacks and equips cybersecurity professionals with the skills to secure AI technologies, moving beyond research hype or business-as-usual strategies. The strategy-based book is a comprehensive guide to AI security, presenting a structured approach with practical examples to identify and counter adversarial attacks. This book goes beyond a random selection of threats and consolidates recent research and industry standards, incorporating taxonomies from MITRE, NIST, and OWASP. Next, a dedicated section introduces a secure-by-design AI strategy with threat

modeling to demonstrate risk-based defenses and strategies, focusing on integrating MLSecOps and LLMOps into security systems. To gain deeper insights, you'll cover examples of incorporating CI, MLOps, and security controls, including open-access LLMs and ML SBOMs. Based on the classic NIST pillars, the book provides a blueprint for maturing enterprise AI security, discussing the role of AI security in safety and ethics as part of Trustworthy AI. By the end of this book, you'll be able to develop, deploy, and secure AI systems effectively.What you will learn Understand poisoning, evasion, and privacy attacks and how to mitigate them Discover how GANs can be used for attacks and deepfakes Explore how LLMs change security, prompt injections, and data exposure Master techniques to poison LLMs with RAG, embeddings, and fine-tuning Explore supply-chain threats and the challenges of open-access LLMs Implement MLSecOps with CIs, MLOps, and SBOMs Who this book is for This book tackles AI security from both angles - offense and defense. AI builders (developers and engineers) will learn how to create secure systems, while cybersecurity professionals, such as security architects, analysts, engineers, ethical hackers, penetration testers, and incident responders will discover methods to combat threats and mitigate risks posed by attackers. The book also provides a secure-by-design approach for leaders to build AI with security in mind. To get the most out of this book, you'll need a basic understanding of security, ML concepts, and Python.

extracting training data from large language models: Hands-On Differential Privacy Ethan Cowan, Michael Shoemate, Mayana Pereira, 2024-05-16 Many organizations today analyze and share large, sensitive datasets about individuals. Whether these datasets cover healthcare details, financial records, or exam scores, it's become more difficult for organizations to protect an individual's information through deidentification, anonymization, and other traditional statistical disclosure limitation techniques. This practical book explains how differential privacy (DP) can help. Authors Ethan Cowan, Michael Shoemate, and Mayana Pereira explain how these techniques enable data scientists, researchers, and programmers to run statistical analyses that hide the contribution of any single individual. You'll dive into basic DP concepts and understand how to use open source tools to create differentially private statistics, explore how to assess the utility/privacy trade-offs, and learn how to integrate differential privacy into workflows. With this book, you'll learn: How DP guarantees privacy when other data anonymization methods don't What preserving individual privacy in a dataset entails How to apply DP in several real-world scenarios and datasets Potential privacy attack methods, including what it means to perform a reidentification attack How to use the OpenDP library in privacy-preserving data releases How to interpret guarantees provided by specific DP data releases

extracting training data from large language models: ECAI 2023 K. Gal, A. Nowé, G.J. Nalepa, 2023-10-18 Artificial intelligence, or AI, now affects the day-to-day life of almost everyone on the planet, and continues to be a perennial hot topic in the news. This book presents the proceedings of ECAI 2023, the 26th European Conference on Artificial Intelligence, and of PAIS 2023, the 12th Conference on Prestigious Applications of Intelligent Systems, held from 30 September to 4 October 2023 and on 3 October 2023 respectively in Kraków, Poland. Since 1974, ECAI has been the premier venue for presenting AI research in Europe, and this annual conference has become the place for researchers and practitioners of AI to discuss the latest trends and challenges in all subfields of AI, and to demonstrate innovative applications and uses of advanced AI technology. ECAI 2023 received 1896 submissions – a record number – of which 1691 were retained for review, ultimately resulting in an acceptance rate of 23%. The 390 papers included here, cover topics including machine learning, natural language processing, multi agent systems, and vision and knowledge representation and reasoning. PAIS 2023 received 17 submissions, of which 10 were accepted after a rigorous review process. Those 10 papers cover topics ranging from fostering better working environments, behavior modeling and citizen science to large language models and neuro-symbolic applications, and are also included here. Presenting a comprehensive overview of current research and developments in AI, the book will be of interest to all those working in the field.

extracting training data from large language models: Statistical Machine Translation Philipp Koehn, 2009-12-17 The dream of automatic language translation is now closer thanks to recent advances in the techniques that underpin statistical machine translation. This class-tested textbook from an active researcher in the field, provides a clear and careful introduction to the latest methods and explains how to build machine translation systems for any two languages. It introduces the subject's building blocks from linguistics and probability, then covers the major models for machine translation: word-based, phrase-based, and tree-based, as well as machine translation evaluation, language modeling, discriminative training and advanced methods to integrate linguistic annotation. The book also reports the latest research, presents the major outstanding challenges, and enables novices as well as experienced researchers to make novel contributions to this exciting area. Ideal for students at undergraduate and graduate level, or for anyone interested in the latest developments in machine translation.

extracting training data from large language models: Security and Privacy in Communication Networks Haixin Duan,

extracting training data from large language models: Explainable AI for Practitioners Michael Munn, David Pitman, 2022-10-31 Most intermediate-level machine learning books focus on how to optimize models by increasing accuracy or decreasing prediction error. But this approach often overlooks the importance of understanding why and how your ML model makes the predictions that it does. Explainability methods provide an essential toolkit for better understanding model behavior, and this practical guide brings together best-in-class techniques for model explainability. Experienced machine learning engineers and data scientists will learn hands-on how these techniques work so that you'll be able to apply these tools more easily in your daily workflow. This essential book provides: A detailed look at some of the most useful and commonly used explainability techniques, highlighting pros and cons to help you choose the best tool for your needs Tips and best practices for implementing these techniques A guide to interacting with explainability and how to avoid common pitfalls The knowledge you need to incorporate explainability in your ML workflow to help build more robust ML systems Advice about explainable AI techniques, including how to apply techniques to models that consume tabular, image, or text data Example implementation code in Python using well-known explainability libraries for models built in Keras and TensorFlow 2.0, PyTorch, and HuggingFace

extracting training data from large language models: Software and Data Engineering Wenying Feng,

extracting training data from large language models: Privacy-Preserving Machine Learning Srinivasa Rao Aravilli, 2024-05-24 Gain hands-on experience in data privacy and privacy-preserving machine learning with open-source ML frameworks, while exploring techniques and algorithms to protect sensitive data from privacy breaches Key Features Understand machine learning privacy risks and employ machine learning algorithms to safeguard data against breaches Develop and deploy privacy-preserving ML pipelines using open-source frameworks Gain insights into confidential computing and its role in countering memory-based data attacks Purchase of the print or Kindle book includes a free PDF eBook Book Description– In an era of evolving privacy regulations, compliance is mandatory for every enterprise – Machine learning engineers face the dual challenge of analyzing vast amounts of data for insights while protecting sensitive information – This book addresses the complexities arising from large data volumes and the scarcity of in-depth privacy-preserving machine learning expertise, and covers a comprehensive range of topics from data privacy and machine learning privacy threats to real-world privacy-preserving cases – As you progress, you'll be guided through developing anti-money laundering solutions using federated learning and differential privacy – Dedicated sections will explore data in-memory attacks and strategies for safeguarding data and ML models – You'll also explore the imperative nature of confidential computation and privacy-preserving machine learning benchmarks, as well as frontier research in the field – Upon completion, you'll possess a thorough understanding of privacy-preserving machine learning, equipping them to effectively shield data from real-world

threats and attacks What you will learn Study data privacy, threats, and attacks across different machine learning phases Explore Uber and Apple cases for applying differential privacy and enhancing data security Discover IID and non-IID data sets as well as data categories Use open-source tools for federated learning (FL) and explore FL algorithms and benchmarks Understand secure multiparty computation with PSI for large data Get up to speed with confidential computation and find out how it helps data in memory attacks Who this book is for – This comprehensive guide is for data scientists, machine learning engineers, and privacy engineers – Prerequisites include a working knowledge of mathematics and basic familiarity with at least one ML framework (TensorFlow, PyTorch, or scikit-learn) – Practical examples will help you elevate your expertise in privacy-preserving machine learning techniques

extracting training data from large language models: Coding with ChatGPT and Other LLMs Dr. Vincent Austin Hall, 2024-11-29 Leverage LLM (large language models) for developing unmatched coding skills, solving complex problems faster, and implementing AI responsibly Key Features Understand the strengths and weaknesses of LLM-powered software for enhancing performance while minimizing potential issues Grasp the ethical considerations, biases, and legal aspects of LLM-generated code for responsible AI usage Boost your coding speed and improve quality with IDE integration Purchase of the print or Kindle book includes a free PDF eBook Book DescriptionKeeping up with the AI revolution and its application in coding can be challenging, but with guidance from AI and ML expert Dr. Vincent Hall—who holds a PhD in machine learning and has extensive experience in licensed software development—this book helps both new and experienced coders to quickly adopt best practices and stay relevant in the field. You'll learn how to use LLMs such as ChatGPT and Bard to produce efficient, explainable, and shareable code and discover techniques to maximize the potential of LLMs. The book focuses on integrated development environments (IDEs) and provides tips to avoid pitfalls, such as bias and unexplainable code, to accelerate your coding speed. You'll master advanced coding applications with LLMs, including refactoring, debugging, and optimization, while examining ethical considerations, biases, and legal implications. You'll also use cutting-edge tools for code generation, architecting, description, and testing to avoid legal hassles while advancing your career. By the end of this book, you'll be well-prepared for future innovations in AI-driven software development, with the ability to anticipate emerging LLM technologies and generate ideas that shape the future of development.What you will learn Utilize LLMs for advanced coding tasks, such as refactoring and optimization Understand how IDEs and LLM tools help coding productivity Master advanced debugging to resolve complex coding issues Identify and avoid common pitfalls in LLM-generated code Explore advanced strategies for code generation, testing, and description Develop practical skills to advance your coding career with LLMs Who this book is for This book is for experienced coders and new developers aiming to master LLMs, data scientists and machine learning engineers looking for advanced techniques for coding with LLMs, and AI enthusiasts exploring ethical and legal implications. Tech professionals will find practical insights for innovation and career growth in this book, while AI consultants and tech hobbyists will discover new methods for training and personal projects.

extracting training data from large language models: *Ubiquitous Security* Guojun Wang,

extracting training data from large language models: Network Simulation and Evaluation Zhaoquan Gu,

extracting training data from large language models: Neural Information Processing Teddy Mantoro, Minho Lee, Media Anugerah Ayu, Kok Wai Wong, Achmad Nizar Hidayanto, 2021-12-06 The two-volume set CCIS 1516 and 1517 constitutes thoroughly refereed short papers presented at the 28th International Conference on Neural Information Processing, ICONIP 2021, held in Sanur, Bali, Indonesia, in December 2021.* The volume also presents papers from the workshop on Artificial Intelligence and Cyber Security, held during the ICONIP 2021. The 176 short and workshop papers presented in this volume were carefully reviewed and selected for publication out of 1093 submissions. The papers are organized in topical sections as follows: theory and algorithms; AI and cybersecurity; cognitive neurosciences; human centred computing; advances in deep and

shallow machine learning algorithms for biomedical data and imaging; reliable, robust, and secure machine learning algorithms; theory and applications of natural computing paradigms; applications. * The conference was held virtually due to the COVID-19 pandemic.

extracting training data from large language models: Introduction to Generative AI Numa Dhamani, Maggie Engler, 2024-02-27 Generative AI tools like ChatGPT are amazing—but how will their use impact our society? This book introduces the world-transforming technology and the strategies you need to use generative AI safely and effectively. Introduction to Generative AI gives you the hows-and-whys of generative AI in accessible language. In this easy-to-read introduction, you'll learn: How large language models (LLMs) work How to integrate generative AI into your personal and professional workflows Balancing innovation and responsibility The social, legal, and policy landscape around generative AI Societal impacts of generative AI Where AI is going Anyone who uses ChatGPT for even a few minutes can tell that it's truly different from other chatbots or question-and-answer tools. Introduction to Generative AI guides you from that first eye-opening interaction to how these powerful tools can transform your personal and professional life. In it, you'll get no-nonsense guidance on generative AI fundamentals to help you understand what these models are (and aren't) capable of, and how you can use them to your greatest advantage. Foreword by Sahar Massachi. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Generative AI tools like ChatGPT, Bing, and Bard have permanently transformed the way we work, learn, and communicate. This delightful book shows you exactly how Generative AI works in plain, jargon-free English, along with the insights you'll need to use it safely and effectively. About the book Introduction to Generative AI guides you through benefits, risks, and limitations of Generative AI technology. You'll discover how AI models learn and think, explore best practices for creating text and graphics, and consider the impact of AI on society, the economy, and the law. Along the way, you'll practice strategies for getting accurate responses and even understand how to handle misuse and security threats. What's inside How large language models work Integrate Generative AI into your daily work Balance innovation and responsibility About the reader For anyone interested in Generative AI. No technical experience required. About the author Numa Dhamani is a natural language processing expert working at the intersection of technology and society. Maggie Engler is an engineer and researcher currently working on safety for large language models. The technical editor on this book was Maris Sekar. Table of Contents 1 Large language models: The power of AI Evolution of natural language processing 2 Training large language models 3 Data privacy and safety with LLMs 4 The evolution of created content 5 Misuse and adversarial attacks 6 Accelerating productivity: Machine-augmented work 7 Making social connections with chatbots 8 What's next for AI and LLMs 9 Broadening the horizon: Exploratory topics in AI

extracting training data from large language models: *Contracting and Contract Law in the Age of Artificial Intelligence* Martin Ebers, Cristina Poncibò, Mimi Zou, 2022-06-30 This book provides original, diverse, and timely insights into the nature, scope, and implications of Artificial Intelligence (AI), especially machine learning and natural language processing, in relation to contracting practices and contract law. The chapters feature unique, critical, and in-depth analysis of a range of topical issues, including how the use of AI in contracting affects key principles of contract law (from formation to remedies), the implications for autonomy, consent, and information asymmetries in contracting, and how AI is shaping contracting practices and the laws relating to specific types of contracts and sectors. The contributors represent an interdisciplinary team of lawyers, computer scientists, economists, political scientists, and linguists from academia, legal practice, policy, and the technology sector. The chapters not only engage with salient theories from different disciplines, but also examine current and potential real-world applications and implications of AI in contracting and explore feasible legal, policy, and technological responses to address the challenges presented by AI in this field. The book covers major common and civil law jurisdictions, including the EU, Italy, Germany, UK, US, and China. It should be read by anyone interested in the complex and fast-evolving relationship between AI, contract law, and related areas of law such as

business, commercial, consumer, competition, and data protection laws.

extracting training data from large language models: Natural Language Processing and Chinese Computing Fei Liu, Nan Duan, Qingting Xu, Yu Hong, 2023-10-07 This three-volume set constitutes the refereed proceedings of the 12th National CCF Conference on Natural Language Processing and Chinese Computing, NLPCC 2023, held in Foshan, China, during October 12–15, 2023. The 143 regular papers included in these proceedings were carefully reviewed and selected from 478 submissions. They were organized in topical sections as follows: dialogue systems; fundamentals of NLP; information extraction and knowledge graph; machine learning for NLP; machine translation and multilinguality; multimodality and explainability; NLP applications and text mining; question answering; large language models; summarization and generation; student workshop; and evaluation workshop.

extracting training data from large language models: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023 Hayit Greenspan, Anant Madabhushi, Parvin Mousavi, Septimiu Salcudean, James Duncan, Tanveer Syeda-Mahmood, Russell Taylor, 2023-09-30 The ten-volume set LNCS 14220, 14221, 14222, 14223, 14224, 14225, 14226, 14227, 14228, and 14229 constitutes the refereed proceedings of the 26th International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2023, which was held in Vancouver, Canada, in October 2023. The 730 revised full papers presented were carefully reviewed and selected from a total of 2250 submissions. The papers are organized in the following topical sections: Part I: Machine learning with limited supervision and machine learning – transfer learning; Part II: Machine learning – learning strategies; machine learning – explainability, bias, and uncertainty; Part III: Machine learning – explainability, bias and uncertainty; image segmentation; Part IV: Image segmentation; Part V: Computer-aided diagnosis; Part VI: Computer-aided diagnosis; computational pathology; Part VII: Clinical applications – abdomen; clinical applications – breast; clinical applications – cardiac; clinical applications – dermatology; clinical applications – fetal imaging; clinical applications – lung; clinical applications – musculoskeletal; clinical applications – oncology; clinical applications – ophthalmology; clinical applications – vascular; Part VIII: Clinical applications – neuroimaging; microscopy; Part IX: Image-guided intervention, surgical planning, and data science; Part X: Image reconstruction and image registration.

extracting training data from large language models: The New Fire Ben Buchanan, Andrew Imbrie, 2022-03-08 AI is revolutionizing the world. Here’s how democracies can come out on top. Artificial intelligence is revolutionizing the modern world. It is ubiquitous—in our homes and offices, in the present and most certainly in the future. Today, we encounter AI as our distant ancestors once encountered fire. If we manage AI well, it will become a force for good, lighting the way to many transformative inventions. If we deploy it thoughtlessly, it will advance beyond our control. If we wield it for destruction, it will fan the flames of a new kind of war, one that holds democracy in the balance. As AI policy experts Ben Buchanan and Andrew Imbrie show in The New Fire, few choices are more urgent—or more fascinating—than how we harness this technology and for what purpose. The new fire has three sparks: data, algorithms, and computing power. These components fuel viral disinformation campaigns, new hacking tools, and military weapons that once seemed like science fiction. To autocrats, AI offers the prospect of centralized control at home and asymmetric advantages in combat. It is easy to assume that democracies, bound by ethical constraints and disjointed in their approach, will be unable to keep up. But such a dystopia is hardly preordained. Combining an incisive understanding of technology with shrewd geopolitical analysis, Buchanan and Imbrie show how AI can work for democracy. With the right approach, technology need not favor tyranny.

extracting training data from large language models: Intelligent Information Processing XII Zhongzhi Shi,

extracting training data from large language models: Deep Network Design for Medical Image Computing Haofu Liao, S. Kevin Zhou, Jiebo Luo, 2022-08-24 Deep Network Design for Medical Image Computing: Principles and Applications covers a range of MIC tasks and discusses

design principles of these tasks for deep learning approaches in medicine. These include skin disease classification, vertebrae identification and localization, cardiac ultrasound image segmentation, 2D/3D medical image registration for intervention, metal artifact reduction, sparse-view artifact reduction, etc. For each topic, the book provides a deep learning-based solution that takes into account the medical or biological aspect of the problem and how the solution addresses a variety of important questions surrounding architecture, the design of deep learning techniques, when to introduce adversarial learning, and more. This book will help graduate students and researchers develop a better understanding of the deep learning design principles for MIC and to apply them to their medical problems. - Explains design principles of deep learning techniques for MIC - Contains cutting-edge deep learning research on MIC - Covers a broad range of MIC tasks, including the classification, detection, segmentation, registration, reconstruction and synthesis of medical images

extracting training data from large language models: Information Security Elias Athanasopoulos, Bart Mennink, 2023-11-30 This book constitutes the proceedings of the 26th International Conference on Information Security, ISC 2023, which took place in Groningen, The Netherlands, in November 2023. The 29 full papers presented in this volume were carefully reviewed and selected from 90 submissions. The contributions were organized in topical sections as follows: privacy; intrusion detection and systems; machine learning; web security; mobile security and trusted execution; post-quantum cryptography; multiparty computation; symmetric cryptography; key management; functional and updatable encryption; and signatures, hashes, and cryptanalysis.

extracting training data from large language models: Generative AI Foundations in Python Carlos Rodriguez, 2024-07-26 Begin your generative AI journey with Python as you explore large language models, understand responsible generative AI practices, and apply your knowledge to real-world applications through guided tutorials Key Features Gain expertise in prompt engineering, LLM fine-tuning, and domain adaptation Use transformers-based LLMs and diffusion models to implement AI applications Discover strategies to optimize model performance, address ethical considerations, and build trust in AI systems Purchase of the print or Kindle book includes a free PDF eBook Book DescriptionThe intricacies and breadth of generative AI (GenAI) and large language models can sometimes eclipse their practical application. It is pivotal to understand the foundational concepts needed to implement generative AI. This guide explains the core concepts behind -of-the-art generative models by combining theory and hands-on application. Generative AI Foundations in Python begins by laying a foundational understanding, presenting the fundamentals of generative LLMs and their historical evolution, while also setting the stage for deeper exploration. You'll also understand how to apply generative LLMs in real-world applications. The book cuts through the complexity and offers actionable guidance on deploying and fine-tuning pre-trained language models with Python. Later, you'll delve into topics such as task-specific fine-tuning, domain adaptation, prompt engineering, quantitative evaluation, and responsible AI, focusing on how to effectively and responsibly use generative LLMs. By the end of this book, you'll be well-versed in applying generative AI capabilities to real-world problems, confidently navigating its enormous potential ethically and responsibly.What you will learn Discover the fundamentals of GenAI and its foundations in NLP Dissect foundational generative architectures including GANs, transformers, and diffusion models Find out how to fine-tune LLMs for specific NLP tasks Understand transfer learning and fine-tuning to facilitate domain adaptation, including fields such as finance Explore prompt engineering, including in-context learning, templatization, and rationalization through chain-of-thought and RAG Implement responsible practices with generative LLMs to minimize bias, toxicity, and other harmful outputs Who this book is for This book is for developers, data scientists, and machine learning engineers embarking on projects driven by generative AI. A general understanding of machine learning and deep learning, as well as some proficiency with Python, is expected.

extracting training data from large language models: Foundation Models for Natural

Language Processing Gerhard Paaß, Sven Giesselbach, 2023-05-23 This open access book provides a comprehensive overview of the state of the art in research and applications of Foundation Models and is intended for readers familiar with basic Natural Language Processing (NLP) concepts. Over the recent years, a revolutionary new paradigm has been developed for training models for NLP. These models are first pre-trained on large collections of text documents to acquire general syntactic knowledge and semantic information. Then, they are fine-tuned for specific tasks, which they can often solve with superhuman accuracy. When the models are large enough, they can be instructed by prompts to solve new tasks without any fine-tuning. Moreover, they can be applied to a wide range of different media and problem domains, ranging from image and video processing to robot control learning. Because they provide a blueprint for solving many tasks in artificial intelligence, they have been called Foundation Models. After a brief introduction to basic NLP models the main pre-trained language models BERT, GPT and sequence-to-sequence transformer are described, as well as the concepts of self-attention and context-sensitive embedding. Then, different approaches to improving these models are discussed, such as expanding the pre-training criteria, increasing the length of input texts, or including extra knowledge. An overview of the best-performing models for about twenty application areas is then presented, e.g., question answering, translation, story generation, dialog systems, generating images from text, etc. For each application area, the strengths and weaknesses of current models are discussed, and an outlook on further developments is given. In addition, links are provided to freely available program code. A concluding chapter summarizes the economic opportunities, mitigation of risks, and potential developments of AI.

extracting training data from large language models: Innovative Technologies and Learning Yueh-Min Huang, Tânia Rocha, 2023 This book constitutes the refereed proceedings of the 6th International Conference on Innovative Technologies and Learning, ICITL 2023, held in Porto, Portugal, during August 28-30, 2023. The 64 full papers included in this book were carefully reviewed and selected from 147 submissions. They cover a wide range of many different research topics, such as: artificial intelligence in education; computational thinking in education; design and framework of learning systems; pedagogies to innovative technologies and learning; STEM/STEAM education; VR/AR/MR/XR in education; and application and design of innovative learning software.

extracting training data from large language models: Clinical Image-Based Procedures, Distributed and Collaborative Learning, Artificial Intelligence for Combating COVID-19 and Secure and Privacy-Preserving Machine Learning Cristina Oyarzun Laura, M. Jorge Cardoso, Michal Rosen-Zvi, Georgios Kaissis, Marius George Linguraru, Raj Shekhar, Stefan Wesarg, Marius Erdt, Klaus Drechsler, Yufei Chen, Shadi Albarqouni, Spyridon Bakas, Bennett Landman, Nicola Rieke, Holger Roth, Xiaoxiao Li, Daguang Xu, Maria Gabrani, Ender Konukoglu, Michal Guindy, Daniel Rueckert, Alexander Ziller, Dmitrii Usynin, Jonathan Passerat-Palmbach, 2021-11-13 This book constitutes the refereed proceedings of the 10th International Workshop on Clinical Image-Based Procedures, CLIP 2021, Second MICCAI Workshop on Distributed and Collaborative Learning, DCL 2021, First MICCAI Workshop, LL-COVID19, First Secure and Privacy-Preserving Machine Learning for Medical Imaging Workshop and Tutorial, PPML 2021, held in conjunction with MICCAI 2021, in October 2021. The workshops were planned to take place in Strasbourg, France, but were held virtually due to the COVID-19 pandemic. CLIP 2021 accepted 9 papers from the 13 submissions received. It focuses on holistic patient models for personalized healthcare with the goal to bring basic research methods closer to the clinical practice. For DCL 2021, 4 papers from 7 submissions were accepted for publication. They deal with machine learning applied to problems where data cannot be stored in centralized databases and information privacy is a priority. LL-COVID19 2021 accepted 2 papers out of 3 submissions dealing with the use of AI models in clinical practice. And for PPML 2021, 2 papers were accepted from a total of 6 submissions, exploring the use of privacy techniques in the medical imaging community.

extracting training data from large language models: Generative AI in Teaching and Learning Hai-Jew, Shalin, 2023-12-05 Generative AI in Teaching and Learning delves into the

revolutionary field of generative artificial intelligence and its impact on education. This comprehensive guide explores the multifaceted applications of generative AI in both formal and informal learning environments, shedding light on the ethical considerations and immense opportunities that arise from its implementation. From the early approaches of utilizing generative AI in teaching to its integration into various facets of learning, this book offers a profound analysis of its potential. Teachers, researchers, instructional designers, developers, data analysts, programmers, and learners alike will find valuable insights into harnessing the power of generative AI for educational purposes.

extracting training data from large language models: Deep Learning Christopher M. Bishop, Hugh Bishop, 2023-11-01 This book offers a comprehensive introduction to the central ideas that underpin deep learning. It is intended both for newcomers to machine learning and for those already experienced in the field. Covering key concepts relating to contemporary architectures and techniques, this essential book equips readers with a robust foundation for potential future specialization. The field of deep learning is undergoing rapid evolution, and therefore this book focusses on ideas that are likely to endure the test of time. The book is organized into numerous bite-sized chapters, each exploring a distinct topic, and the narrative follows a linear progression, with each chapter building upon content from its predecessors. This structure is well-suited to teaching a two-semester undergraduate or postgraduate machine learning course, while remaining equally relevant to those engaged in active research or in self-study. A full understanding of machine learning requires some mathematical background and so the book includes a self-contained introduction to probability theory. However, the focus of the book is on conveying a clear understanding of ideas, with emphasis on the real-world practical value of techniques rather than on abstract theory. Complex concepts are therefore presented from multiple complementary perspectives including textual descriptions, diagrams, mathematical formulae, and pseudo-code. Chris Bishop is a Technical Fellow at Microsoft and is the Director of Microsoft Research AI4Science. He is a Fellow of Darwin College Cambridge, a Fellow of the Royal Academy of Engineering, and a Fellow of the Royal Society. Hugh Bishop is an Applied Scientist at Wayve, a deep learning autonomous driving company in London, where he designs and trains deep neural networks. He completed his MPhil in Machine Learning and Machine Intelligence at Cambridge University. “Chris Bishop wrote a terrific textbook on neural networks in 1995 and has a deep knowledge of the field and its core ideas. His many years of experience in explaining neural networks have made him extremely skillful at presenting complicated ideas in the simplest possible way and it is a delight to see these skills applied to the revolutionary new developments in the field.” -- Geoffrey Hinton With the recent explosion of deep learning and AI as a research topic, and the quickly growing importance of AI applications, a modern textbook on the topic was badly needed. The New Bishop masterfully fills the gap, covering algorithms for supervised and unsupervised learning, modern deep learning architecture families, as well as how to apply all of this to various application areas. – Yann LeCun “This excellent and very educational book will bring the reader up to date with the main concepts and advances in deep learning with a solid anchoring in probability. These concepts are powering current industrial AI systems and are likely to form the basis of further advances towards artificial general intelligence.” -- Yoshua Bengio

extracting training data from large language models: Natural Language Processing and Chinese Computing Derek F. Wong,

extracting training data from large language models: Generative AI for Web Engineering Models Shah, Imdad Ali, Jhanjhi, Noor Zaman, 2024-10-22 Web engineering faces a pressing challenge in keeping pace with the rapidly evolving digital landscape. Developing, designing, testing, and maintaining web-based systems and applications require innovative approaches to meet the growing demands of users and businesses. Generative Artificial Intelligence (AI) emerges as a transformative solution, offering advanced capabilities to enhance web engineering models and methodologies. This book presents a timely exploration of how Generative AI can revolutionize the web engineering discipline, providing insights into future challenges and societal impacts.

Generative AI for Web Engineering Models offers a comprehensive examination of integrating AI-driven generative approaches into web engineering practices. It delves into methodologies, models, and the transformative impact of Generative AI on web-based systems and applications. By addressing topics such as web browser technologies, website scalability, security, and the integration of Machine Learning, this book provides a roadmap for researchers, scientists, postgraduate students, and AI enthusiasts interested in the intersection of AI and web engineering.

extracting training data from large language models: Representation Learning for Natural Language Processing Zhiyuan Liu, Yankai Lin, Maosong Sun, 2023-08-23 This book provides an overview of the recent advances in representation learning theory, algorithms, and applications for natural language processing (NLP), ranging from word embeddings to pre-trained language models. It is divided into four parts. Part I presents the representation learning techniques for multiple language entries, including words, sentences and documents, as well as pre-training techniques. Part II then introduces the related representation techniques to NLP, including graphs, cross-modal entries, and robustness. Part III then introduces the representation techniques for the knowledge that are closely related to NLP, including entity-based world knowledge, sememe-based linguistic knowledge, legal domain knowledge and biomedical domain knowledge. Lastly, Part IV discusses the remaining challenges and future research directions. The theories and algorithms of representation learning presented can also benefit other related domains such as machine learning, social network analysis, semantic Web, information retrieval, data mining and computational biology. This book is intended for advanced undergraduate and graduate students, post-doctoral fellows, researchers, lecturers, and industrial engineers, as well as anyone interested in representation learning and natural language processing. As compared to the first edition, the second edition (1) provides a more detailed introduction to representation learning in Chapter 1; (2) adds four new chapters to introduce pre-trained language models, robust representation learning, legal knowledge representation learning and biomedical knowledge representation learning; (3) updates recent advances in representation learning in all chapters; and (4) corrects some errors in the first edition. The new contents will be approximately 50%+ compared to the first edition. This is an open access book.

extracting training data from large language models: Ethics and Fairness in Medical Imaging Esther Puyol-Antón,

extracting training data from large language models: Data Science and Artificial Intelligence Chutiporn Anutariya, Marcello M. Bonsangue, 2023-11-17 This book constitutes the proceedings of the First International Conference, DSAI 2023, held in Bangkok, Thailand, during November 27–30, 2023. The 22 full papers and the 4 short papers included in this volume were carefully reviewed and selected from 70 submissions. This volume focuses on ideas, methodologies, and cutting-edge research that can drive progress and foster interdisciplinary collaboration in the fields of data science and artificial intelligence.

extracting training data from large language models: Distributed Computer and Communication Networks Vladimir M. Vishnevskiy, Dmitry V. Kozyrev, Konstantin E. Samouylov, 2024 Zusammenfassung: This book constitutes the refereed proceedings of the 26th International Conference on Distributed Computer and Communication Networks: Control, Computation, Communications, DCCN 2023, held in Moscow, Russia, during September 25-29, 2023. The 37 full papers and 4 short papers included in this book were carefully reviewed and selected from 122 submissions. They were organized in topical sections as follows: Distributed Systems Applications; Analytical Modeling of Distributed Systems; Computer and Communication Networks

extracting training data from large language models: Proceedings of the NIELIT's International Conference on Communication, Electronics and Digital Technology Palaiahnakote Shivakumara,

extracting training data from large language models: AI Doctor Ronald M. Razmi, MD, 2024-01-31 Explores the transformative impact of artificial intelligence (AI) on the healthcare industry AI Doctor: The Rise of Artificial Intelligence in Healthcare provides a timely and

authoritative overview of the current impact and future potential of AI technology in healthcare. With a reader-friendly narrative style, this comprehensive guide traces the evolution of AI in healthcare, describes methodological breakthroughs, drivers and barriers of its adoption, discusses use cases across clinical medicine, administration and operations, and life sciences, and examines the business models for the entrepreneurs, investors, and customers. Detailed yet accessible chapters help those in the business and practice of healthcare recognize the remarkable potential of AI in areas such as drug discovery and development, diagnostics, therapeutics, clinical workflows, personalized medicine, early disease prediction, population health management, and healthcare administration and operations. Throughout the text, author Ronald M. Razmi, MD offers valuable insights on harnessing AI to improve health of the world population, develop more efficient business models, accelerate long-term economic growth, and optimize healthcare budgets. Addressing the potential impact of AI on the clinical practice of medicine, the business of healthcare, and opportunities for investors, AI Doctor: The Rise of Artificial Intelligence in Healthcare: Discusses what AI is currently doing in healthcare and its direction in the next decade Examines the development and challenges for medical algorithms Identifies the applications of AI in diagnostics, therapeutics, population health, clinical workflows, administration and operations, discovery and development of new clinical paradigms and more Presents timely and relevant information on rapidly expanding generative AI technologies, such as Chat GPT Describes the analysis that needs to be made by entrepreneurs and investors as they evaluate building or investing in health AI solutions Features a wealth of relatable real-world examples that bring technical concepts to life Explains the role of AI in the development of vaccines, diagnostics, and therapeutics during the COVID-19 pandemic AI Doctor: The Rise of Artificial Intelligence in Healthcare. A Guide for Users, Buyers, Builders, and Investors is a must-read for healthcare professionals, researchers, investors, entrepreneurs, medical and nursing students, and those building or designing systems for the commercial marketplace. The book's non-technical and reader-friendly narrative style also makes it an ideal read for everyone interested in learning about how AI will improve health and healthcare in the coming decades.

extracting training data from large language models: Bridging the Gap Between AI and Reality Bernhard Steffen,

extracting training data from large language models: *The Future of Digital-Physical Interactions* Siva Sathyanarayana Movva, Siva Karthik Devineni, Moses Michael Meitivyeki, Ankur Tak, Kodanda Rami Reddy Manukonda, 2024-06-06 TOPICS IN THE BOOK Development of Digital Twins for Urban Water Systems AI-Enhanced Data Visualization: Transforming Complex Data into Actionable Insights Global Positioning System Signal Verification through Correlation Function Distortion and Received Power Tracking Risk Management in Agile Al/Ml Projects: Identifying and Mitigating Data and Model Risks Addressing Challenges in Test Automation Adoption: A Study on Strategies for Overcoming Barriers to Seamless QA Integration

extracting training data from large language models: Legal Knowledge and Information Systems G. Sileno, J. Spanakis, G. van Dijck, 2023-12-19 Technological advances related to legal information, knowledge representation, engineering, and processing have aroused growing interest within the research community and the legal industry in recent years. These advances relate to areas such as computational and formal models of legal reasoning, legal data analytics, legal information retrieval, the application of machine learning techniques to different legal tasks, and the experimental evaluation of these systems. This book presents the proceedings of JURIX 2023, the 36th International Conference on Legal Knowledge and Information Systems, held from 18–20 December 2023 in Maastricht, the Netherlands. This annual conference has become recognized as an international forum where academics and professionals working at the intersection of law and artificial intelligence can exchange knowledge and experience. A total of 92 submissions were received for the conference, of which 18 were selected as long papers, 30 as short papers and 7 as demo papers following a rigorous review process. This represents an acceptance rate of around 20% for long papers (60% overall). Topics covered include formal approaches applied to various aspects

of legal reasoning; machine learning and information retrieval methods applied to various natural language processing tasks; hybrid approaches to working on the frontier between symbolic and sub-symbolic methods; experimental inquiries into the interfaces between computational systems and legal systems; and network analysis in law. Providing a comprehensive overview of recent advances in the field, the book will be of interest to all those working at the intersection between law and AI.

EXTRACTING | English meaning - Cambridge Dictionary
EXTRACTING definition: 1. present participle of extract 2. to remove or take out something: 3. to make someone give you…. Learn more.

EXTRACT Definition & Meaning - Merriam-Webster
Extract forms a kind of mirror image of abstract: more common as a verb, but also used as a noun and adjective. The …

Extracting - definition of extracting by The Free Dictionary
Define extracting. extracting synonyms, extracting pronunciation, extracting translation, English dictionary definition of …

Extract Definition & Meaning | Britannica Dictionary
Investigators were able to extract useful information from the company's financial records. They are hoping to extract new …

EXTRACT definition and meaning | Collins English Dictionary
To extract a substance means to obtain it from something else, for example by using industrial or chemical processes. ...the …

extracting - WordReference.com Dictionary of English
To extract is to draw forth something as by pulling, importuning, or the like: to extract a confession by torture. To exact is to …

extract verb - Definition, pictures, pronunciation and usage notes ...
extract something (from something) to remove or obtain a substance from something, for example by using an …

EXTRACTION Definition & Meaning - Merriam-Webster
The meaning of EXTRACTION is the act or process of extracting something. How to use extraction in a sentence.

What does extracting mean? - Definitions.net
An extract is a substance made by extracting a part of a raw material, often by using a solvent such as ethanol, oil or water. …

EXTRACT | English meaning - Cambridge Dictionary
The science of extracting useful information from large data sets is usually referred to as 'data mining', sometimes along with …

EXTRACTING | English meaning - Cambridge Dictionary
EXTRACTING definition: 1. present participle of extract 2. to remove or take out something: 3. to make someone give you…. Learn more.

EXTRACT Definition & Meaning - Merriam-Webster
Extract forms a kind of mirror image of abstract: more common as a verb, but also used as a noun and adjective. The adjective, meaning …

Extracting - definition of extracting by The Free Dictionary
Define extracting. extracting synonyms, extracting pronunciation, extracting translation, English dictionary definition of extracting. …

Extract Definition & Meaning | Britannica Dictionary
Investigators were able to extract useful information from the company's financial records. They are hoping to extract new …

EXTRACT definition and meaning | Collins English Dictionary
To extract a substance means to obtain it from something else, for example by using industrial or chemical processes. …the traditional …

Back to Home