20 recommended free AI-Ethics, Data Ethics and XAI online courses to get started right away[Updated 13/09/22]

Murat Durmus (CEO @AISOMA_AG) · Follow
Published in Nerd For Tech
17 min read · Jul 8, 2021

Listen Share

(shutterstock: hafakot)

AI ethics is becoming more and more important. Discussions about AI Ethics are still mostly conducted in academic circles. But one can already see that many companies are seriously dealing with it. One thing that seems clear to me:

> **Graduates of philosophy and ethics will be in high demand in the future to investigate AI-related processes through a human lens.**

(taken from the article: Do Companies need a Chief AI-Ethics Officer?)

Contents:

1. Ethics of AI (University of Helsinki)

The Ethics of AI is a free online course created by the University of Helsinki. The course is for anyone who is interested in the ethical aspects of AI — we want to

encourage people to learn what AI ethics means, what can and can't be done to develop AI in an ethically sustainable way, and how to start thinking about AI from an ethical point of view.

Content:

Chapter 1: What is AI ethics? - What does AI ethics mean and what role do values and norms play? We'll also look at the principles of AI ethics that we will follow in this course.

Chapter 2: What should we do? - What do the principles of beneficence (do good) and non-maleficence (do no harm) mean for AI, and how do they relate to the concept of the "common good?

Chapter 3: Who should be blamed? - What does accountability actually mean, and how does it apply to AI ethics? We'll also discuss what moral agency and responsibility mean and the difficulty of assigning blame.

Chapter 4: Should we know how AI works - Why is transparency in AI important and what major issues are affected by transparency — and what are some of the risks associated with transparency in AI systems?

Chapter 5: Should AI respect and promote rights? - What are human rights, and how do they tie into the current ethical guidelines and principles of AI? We'll also look more closely at three rights of particular importance to AI: the right to privacy, security, and inclusion.

Chapter 6: Should AI be fair and non-discriminative - What does fairness mean in relation to AI, how does discrimination manifest through AI — and what can we do to make these systems less biased?

Chapter 7: AI ethics in practice - What are some of the current challenges for AI ethics, what role do AI guidelines play in shaping the discussion, and how might things develop in the future?

Course: **Ethics of AI**

2. AI-Ethics: Global Perspectives (aiethicscourse.org)

Data-intensive and AI-based technologies can solve the world's biggest challenges, but they also pose risks to individuals and groups. As we deploy new technology, we must consider ethical ramifications of AI use to identify and rectify harms.

This course is designed to raise awareness of the societal impacts of technology and to give individuals and institutions the tools to pursue responsible AI use. Intended for current and future data scientists, policymakers, and business leaders, this course contains **19 modules** on topics related to artificial intelligence. Each module consists of a video lecture accompanied by additional resources such as podcasts, videos, and readings.

Course: **AI Ethics: Global Perspectives**

3. AI Ethics for Business (Seattle University)

This course is meant as an introduction to the ethical dimension of the uses of AI technologies. Upon successful completion of the course, students will 1) be sensitive to ethical issues surrounding transformative technologies, and 2) be able to articulate possible courses of action when there are ethically sensitive issues.

Mod 0 — Course Orientation

Mod 1 — Values and Moral Excuses in AI Technologies

Mod 2 — Transformative Technologies and their Impacts

Mod 3 — Users of Technology

Time commitment: 6–10 hours

Course: **AI Ethics for Business**

4. Bias and Discrimination in AI (Université de Montréal)

Discover how even computer algorithms may be biased and have a serious impact on our every day lives. In this MOOC, based on an IVADO School involving various international experts in the field, you will learn how to identify and alleviate bias and discrimination in Artificial Intelligence.

Engage in this course pertaining to a highly impactful yet, too rarely discussed, AI-related topic. You will learn from international experts in the field, also speakers at IVADO's International School on Bias and Discrimination in AI, which took place in Montreal, and explore the social and technical aspects of bias, discrimination and fairness in machine learning and algorithm design.

The main focus of this course is: gender, race and socioeconomic-based bias as well as bias in data-driven predictive models leading to decisions. The course is primarily intended for professionals and academics with basic knowledge in mathematics and programming, but the rich content will be of great use to whomever uses, or is interested in, AI in any other way. These sociotechnical topics have proven to be great eye-openers for technical professionals! (Université de Montréal)

What you'll learn

- Understanding bias and discrimination in all its aspects
- Exploring the harmful effects of bias in machine learning (discriminatory effects of algorithmic decision-making)
- Identifying the sources of bias and discrimination in machine learning
- Mitigating bias in machine learning (strategies for addressing bias)
- Recommendations to guide the ethical development and evaluation of algorithms

Course: **Bias and Discrimination in AI**

5. Data Science Ethics (University of Michigan)

Learn how to think through the ethics surrounding privacy, data sharing, and algorithmic decision-making.

This course focused on ethics specifically related to data science will provide you with the framework to analyze these concerns. This framework is based on ethics, which are shared values that help differentiate right from wrong. Ethics are not law, but they are usually the basis for laws.

Everyone, including data scientists, will benefit from this course. No previous knowledge is needed.

What you'll learn

- Who owns data
- How we value different aspects of privacy
- How we get informed consent
- What it means to be fair

Course: Data Science Ethics

6. Intro to AI Ethics (Kaggle)

Explore practical tools to guide the moral design of AI systems by Alexis Cook and Var Shankar

1. Introduction to AI Ethics
Learn what to expect from the course.

2. Human-Centered Design for AI
Design systems that serve people's needs. Navigate issues in several real-world scenarios.

3. Identifying Bias in AI
Bias can creep in at any stage in the pipeline. Investigate a simple model that identifies toxic text.

4. AI Fairness
Learn about four different types of fairness. Assess a toy model trained to judge credit card applications.

5. Model Cards
Increase transparency by communicating key information about machine learning models.

Course: https://www.kaggle.com/learn/intro-to-ai-ethics

7. Ethics in AI and Data Science (LFS112x)

Learn how to build and incorporate ethical principles and frameworks in your AI and Data Science technology and business initiatives to add transparency, build trust, drive

adoption, and lead with responsibility.

Who Is It For

The course is designed for business, government and technology leaders and data scientists who are responsible for building and adopting AI tools.

What You'll Learn

In this course you will learn about business drivers for AI, the ethical challenges and impacts of AI and Data Science, the business and societal dynamics at work in an AI world, the key principles for building responsible AI, and more. This course introduces some of the principles and frameworks that puts ethics and responsibility into practice in the data analytics profession. And offers practical approaches to technical, business and leadership dilemmas and challenges posed by work in AI and Data Science.

What It Prepares You For

You will walk away from this course with an understanding of how to add transparency, develop standards and share best practices to build trust and drive AI adoption.

Course: https://training.linuxfoundation.org/training/ethics-in-ai-and-data-science-lfs112/

8. Practical Data Ethics (Fast AI)

This class was originally taught in-person at the University of San Francisco Data Institute in January-February 2020, for a diverse mix of working professionals from a range of backgrounds (as an evening certificate courses). There are no prerequisites for the course. This course is in no way intended to be exhaustive, but hopefully will provide useful context about how data misuse is impacting society, as well as practice in critical thinking skills and questions to ask.

Topics covered:

1. Disinformation
2. Bias & Fairness

3. Ethical Foundations & Practical Tools

4. Privacy & surveillance

5. Our Ecosystem: Metrics, Venture Capital, & Losing the Forest for the Trees

6. Algorithmic Colonialism, and Next Steps

Desired Learning Outcomes

1. Understand the impacts of data misuse, including unjust bias, surveillance, disinformation, and feedback loops. Understand the contributing factors to these impacts. Identify different types of bias.

2. Develop literacy in investigating how data and data-powered algorithms shape, constrain, and manipulate our commercial, civic, and personal experiences.

3. Analyze new scenarios and potential products to try to identify and mitigate potential risks.

4. Have a toolkit of ethical techniques and practices to implement in their workplaces

Check out the syllabus and reading list!

Watch the videos here!

9. Data Ethics, AI and Responsible Innovation (University of Edinburgh)

About this course

Skip About this course

How much would you like your smart home to know about you? Has your data been harvested and used for political advertising on social media? Would you be happy to be profiled by a predictive policing AI?

What you'll learn

· Understand and articulate the critical, social, legal, political and ethical issues arising throughout the data lifecycle.

· Understand relevant concepts, including: ethics/morality, responsibility, digital rights, data governance, human-data interaction, responsible research and innovation.

· Identify and assess current ethical issues in data science and industry.

· Apply professional critical judgement and reflexivity to moral problems with no clear solutions.

· Evaluate ethical issues you face in your current professional practice.

· Identify and apply ethically driven solutions to those issues.

Course: https://www.edx.org/course/Data-Ethics-AI-and-Responsible-Innovation

10. Identify guiding principles for responsible AI (Microsoft)

At Microsoft, we've recognized six principles that we believe should guide AI development and use — fairness, reliability and safety, privacy and security, inclusiveness, transparency, and accountability. For us, these principles are the cornerstone of a responsible and trustworthy approach to AI, especially as intelligent technology becomes more prevalent in the products and services we use every day. We recognize that every individual, company, and region will have their own beliefs and standards that should be reflected in their AI journey. We want to share our perspective as you consider developing your own guiding principles.

Learning objectives

In this module, you will:

· Describe the importance of engaging with AI in a responsible manner.

· Identify six guiding principles to develop and use AI responsibly.

· Describe different approaches to responsible AI governance from experienced executives.

Course:

https://docs.microsoft.com/en-us/learn/modules/responsible-ai-principles/

11. Human-Computer Interaction III: Ethics, Needfinding & Prototyping (Georgia Tech)

About this course

This course takes you through lessons 14 through 18 of CS6750: Human-Computer Interaction as taught in the Georgia Tech Online Master of Science in Computer Science program.

In this course, you'll begin by learning the design life cycle. This is the process by which we investigate user needs, brainstorm potential designs, create prototypes, and evaluate those prototypes. This life cycle provides the structure for the third and fourth courses in this professional certificate.

What you'll learn

· The structure of the design life cycle: needfinding, brainstorming, prototyping, and evaluation.

· The value of research ethics and the importance of emphasizing the user's rights.

· The role of Institutional Review Boards in governing university research.

· The importance of ethics in industry and the mechanisms for ensuring they are maintained.

· The importance of needfinding in the design life cycle.

· Mechanisms for active needfinding, including surveys, interviews, and focus groups.

· Approaches to personal needfinding, including participant observation and apprenticeship.

· Ways to observe organic interaction, such as naturalistic observation or investigation of hacks and workarounds.

· The data inventory, a structure for ensuring you understand your user and their problems.

· Approaches to brainstorming design ideas, both individually and in groups.

· Methods for further exploring design ideas, such as user personas and storyboards.

· The spectrum of prototyping, from low-fidelity to high-fidelity.

· The importance of low-fidelity prototypes in getting early feedback.

Course: Human-Computer Interaction III: Ethics, Needfinding & Prototyping

12. Ethics in Action

What do the world's great religious and secular philosophies have to say about ethical conduct? Which virtues are common across faiths? And what role do religious communities have to play in building a more just and sustainable world?

About this course

The challenges of sustainable development are more than technical or political — they are also moral, calling on us to examine who we are as human beings, and who we want to be going forward.

This examination of what makes a "good person" — called virtue ethics — has long been the purview of philosophers and theologians. So what do the world's great religious and secular philosophies have to say about ethical conduct? Which virtues are common across creeds and cultures? And what role do ethics, spirituality, and religious communities have to play in sustainable development?

In 2016, distinguished leaders from the world's major religious traditions, philosophers, scholars, and scientists were invited by Pope Francis to the Vatican for a series of meetings. Known as the Ethics in Action initiative, these meetings sought to promote dialogue and find consensus about the values needed to advance transformative action for our common home and the human family.

This course features Ethics in Action meeting participants, as well as other leading voices, and discusses the perspectives of some of the world's great traditions and their role in addressing our world's most pressing challenges, including poverty, corruption, and climate change. It presents multi-faith cooperation as essential for achieving

sustainable development, and calls for the development of a new shared virtue ethics to bring us all into a sustainable and peaceful future.

This course is for:

Religious and spiritual communities and individuals who want to contribute to the work of sustainable development

Development professionals who want to understand and engage local faith communities

Advanced undergraduates and graduate students interested in philosophy, religion, theology, and their relation to global issues

Course: **Ethics in Action**

13. Explainable Machine Learning with LIME and H2O in R (coursera)

Welcome to this hands-on, guided introduction to Explainable Machine Learning with LIME and H2O in R. By the end of this project, you will be able to use the LIME and H2O packages in R for automatic and interpretable machine learning, build classification models quickly with H2O AutoML and explain and interpret model predictions using LIME.

About this course

1. Introduction and Project Overview

2. Import Libraries and Load the IBM HR Employee Attrition Data

3. Preprocess Data using Recipes

4. Start H2O Cluster and Create Train/Test Splits

5. Run AutoML to Train and Tune Models

6. Leaderboard Exploration

7. Model Performance Evaluation

8. Local Interpretable Model-Agnostic Explanations (LIME)

9. Apply LIME to Interpret Model Outcomes

Course: Explainable Machine Learning with LIME and H2O in R

14. An introduction to explainable AI, and why we need it

This is an online tutorial. You will be exposed to a brief introduction to explainable AI, how it works and its importance. The blog will help you understand the Reversed Time Attention Model (RETAIN) model, Local Interpretable Model-Agnostic Explanations (LIME), how explainable AI works as one generates newer and more innovative applications for neural networks. Author Patrick Ferris explainS all of these through instances such as the one-pixel attack.

Course: An introduction to explainable AI, and why we need it

15. Explainable AI: Scene Classification and GradCam Visualization (coursera)

In this 2 hour long hands-on project, we will train a deep learning model to predict the type of scenery in images. In addition, we are going to use a technique known as Grad-Cam to help explain how AI models think. This project could be practically used for detecting the type of scenery from the satellite images.

About this course

1. Understand the theory and intuition behind Deep Neural Networks, Residual Nets, and Convolutional Neural Networks (CNNs)

2. Apply Python libraries to import, pre-process and visualize images

3. Perform data augmentation to improve model generalization capability

4. Build a deep learning model based on Convolutional Neural Network and Residual blocks using Keras with Tensorflow 2.0 as a backend

5. Compile and fit Deep Learning model to training data

6. Assess the performance of trained CNN and ensure its generalization using various KPIs such as accuracy, precision and recall

7. Understand the theory and intuition behind GradCam and Explainable AI

8. Visualize the Activation Maps used by CNN to make predictions using Grad-CAM

Course: Explainable AI: Scene Classification and GradCam Visualization

16. Interpretable Machine Learning Applications: Part 1 & 2 (coursera)

Part1: In this 1-hour long project-based course, you will learn how to create interpretable machine learning applications on the example of two classification regression models, decision tree and random forestc classifiers. You will also learn how to explain such prediction models by extracting the most important features and their values, which mostly impact these prediction models. In this sense, the project will boost your career as Machine Learning (ML) developer and modeler in that you will be able to get a deeper insight into the behaviour of your ML model. The project will also benefit your career as a decision maker in an executive position, or consultant, interested in deploying trusted and accountable ML applications.

About this course

1. Setting the stage (Python Jupyter Lab web-based Server environment, importing the dataset and file to train and test the designated classification regressors as prediction models).

2. Train, test and estimate the accuracy (confusion matrix) of a Decision Tree classifier.

3. Train, test and estimate the accuracy (confusion matrix) of a Random Tree classifier as an alternative to the previous one.

4. Extract a ranking list of the features, which are most important for each one of our prediction models.

5. Extract and plot the impact of the values of selected important features on predictions being made by each one of our prediction models.

Part2: By the end of this project, you will be able to develop intepretable machine learning applications explaining individual predictions rather than explaining the behavior of the prediction model as a whole. This will be done via the well known Local Interpretable Model-agnostic Explanations (LIME) as a machine learning interpretation and explanation model. In particular, in this project, you will learn how to go beyond the development and use of machine learning (ML) models, such as regression classifiers, in that we add on explainability and interpretation aspects for individual predictions.

About this course

1. Explore and understand the features and values from the available data about red wine quality

2. Transform the available data into a classification dataset and problem

3. Prepare the data for training and validation purposes

4. Train, validate, estimate, and contrast the performance of three regression classifiers: Decision Tree, Random Forest, AdaBoost

5. Prepare and train the “explainer” in terms of the LIME library

6. Display and interpret explanations of individual predictions made by the three classifiers

Part 1

Part 2

17. Foundations of Ethical Decision-Making (Georgetown University)

About this course

Skip About this course

Globalization has dramatically increased the influence of international corporations in political, economic, and sociocultural spheres. More than ever, global managers are faced with profound choices about their impact. In this course, you will analyze real-

world ethical dilemmas using multiple frameworks and ultimately refining your own approach to ethical decision-making.

When considering complex ethical questions, how useful are your starting assumptions? How do your decisions change when you prioritize established rules versus projected outcomes? Through case studies on mineral extraction and racial discrimination, you will gain a practical foundation in applied ethics and the skills to makes sound ethical decisions throughout your career.

What you'll learn

· Describe the characteristics of applied ethics and the impact of ethical decision-making on society.

· Define and contrast rules- and results-based approaches to ethical decision-making.

· Apply rules-based and results-based approaches in relation to human rights scenarios.

· Analyze the role of corporate political involvement in balancing human rights with the recognition of host governments' legitimacy.

· Identify and assess how global business practices intersect with values of censorship, privacy, and public safety on the internet.

Course: Foundations of Ethical Decision-Making: Government and Political Issues

18. Philosophy and Critical Thinking

About this course

Skip About this course

What can we learn through philosophical inquiry that will help us to think with clarity, rigour and humour about things that matter?

This course introduces principles of philosophical inquiry and critical thinking that will help us answer this question. Learn how we can use philosophical ideas to think about ourselves and the world around us.

What you'll learn

· How to think with clarity and rigour

· How to identify, analyse and construct cogent arguments

· How to think of solutions tothe central problems of philosophy

· How to engage in philosophical conversationswith others about topics that matter

Course: Philosophy and Critical Thinking

19. IEEE Awareness Module on AI Ethics

AI ethics is central to the adoption of AI. Regulations are being developed across the world that will determine ethicality and market access. This course will explain to decision makers how ethicality is assessed in an AI solution, the differences between the risk-based approaches, and the advantages of outcomes based risk models. It will also introduce IEEE CertifAIEd AI ethics evaluation solution for AIS, how it works, and how to test an AI solution for ethicality with IEEE CertifAIEd.

Pre -requisites

· No specific prerequisites

· General understanding of products development and deployment

· General understanding of certification and regulatory processes

Course: IEEE Awareness Module on AI Ethics

20. Introduction to Corporate Sustainability, Social Innovation and Ethics

About this course

Skip About this course

Sustainability is one of the key issuesfacing today's society. This is underlined by the increasing attention on sustainability issues by governments, media, academics and industry.

In the context of sustainable development, businesses that are often referredto as part of the problem can be part of the solution. As a consequence, policymakers, industry leaders, society and academics are trying to understand how sustainability affects traditional ways of doing business, and also, how traditional businesses are affected by sustainability. How to develop a sustainable competitive advantage is a key challenge inthe agendas of today's global executives.

This MBA Primer introduces the topic of sustainability from three perspectives: corporate sustainability (core of the module); business ethics and social innovation.

The virtual class will be involved in an engaging debatediscussing the different economic, environmental and social perspectives on the topics discussed. At the end of this module, participants will have a comprehensive understanding of sustainability issues, the relevance for policymakers, the role of corporations and the implications for decision-making.

What you'll learn

Participants on this module can expect to:

· Develop an understanding of the strategic challenges faced by modern organisations in today's global business environment

· Develop a basic understanding of sustainability thinking and triple-bottom-line approaches

· Develop an appreciation of the business case for sustainability in different industries

· Develop an appreciation of the ethical implications of business decisions

· Develop an appreciation of the potential for social innovation through business activities.

In this module, we will work through severalcase studies and hear from a range of speakers from industry and academia.

Course: Introduction to Corporate Sustainability, Social Innovation and Ethics

This might be also of interest "*THE AI THOUGHT BOOK*" (my latest book on 'Mindful AI')

More Info and free excerpt to download: ***THE AI THOUGHT BOOK***

Artificial Intelligence | Education | Ethics | Ai Ethics | Course

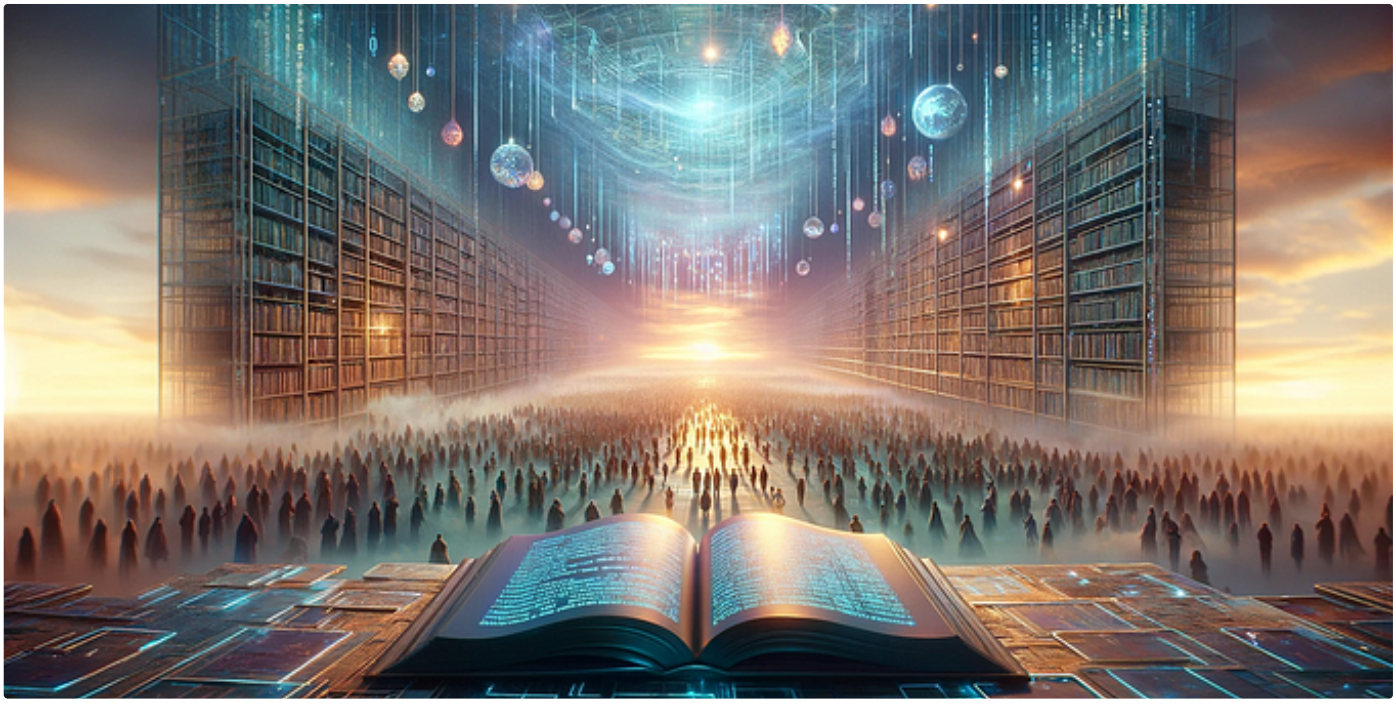
Follow

Written by Murat Durmus (CEO @AISOMA_AG)

514 Followers · Writer for Nerd For Tech

CEO & Founder @AISOMA_AG | Author | #ArtificialIntelligence | #CEO | #AI | #AIStrategy | #Leadership | #Philosophy | #AIEthics | (views are my own)

More from Murat Durmus (CEO @AISOMA_AG) and Nerd For Tech



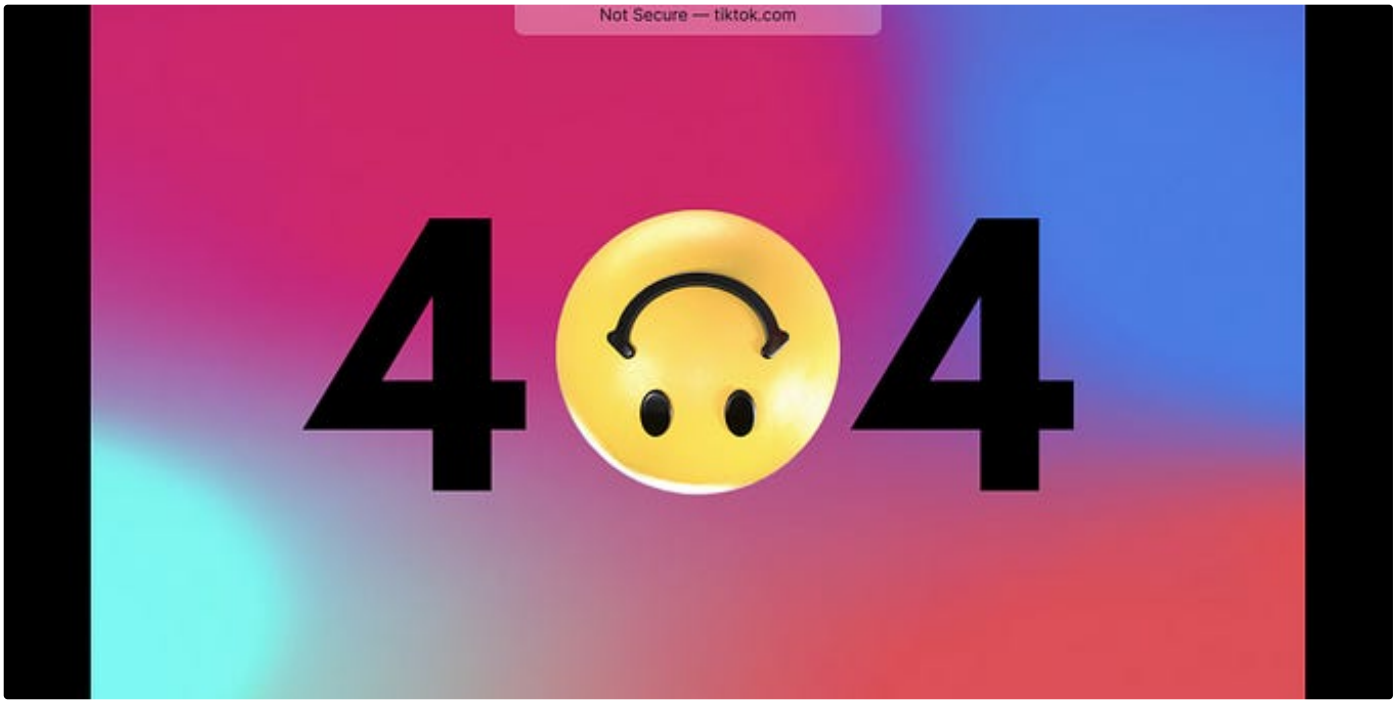
Murat Durmus (CEO @AISOMA_AG)

If ChatGPT had a Worldview

If ChatGPT did have a worldview, it would be a curious amalgam of relentless objectivity and a chameleon-like adaptability. Picture a vast...

2 min read · Nov 15

19



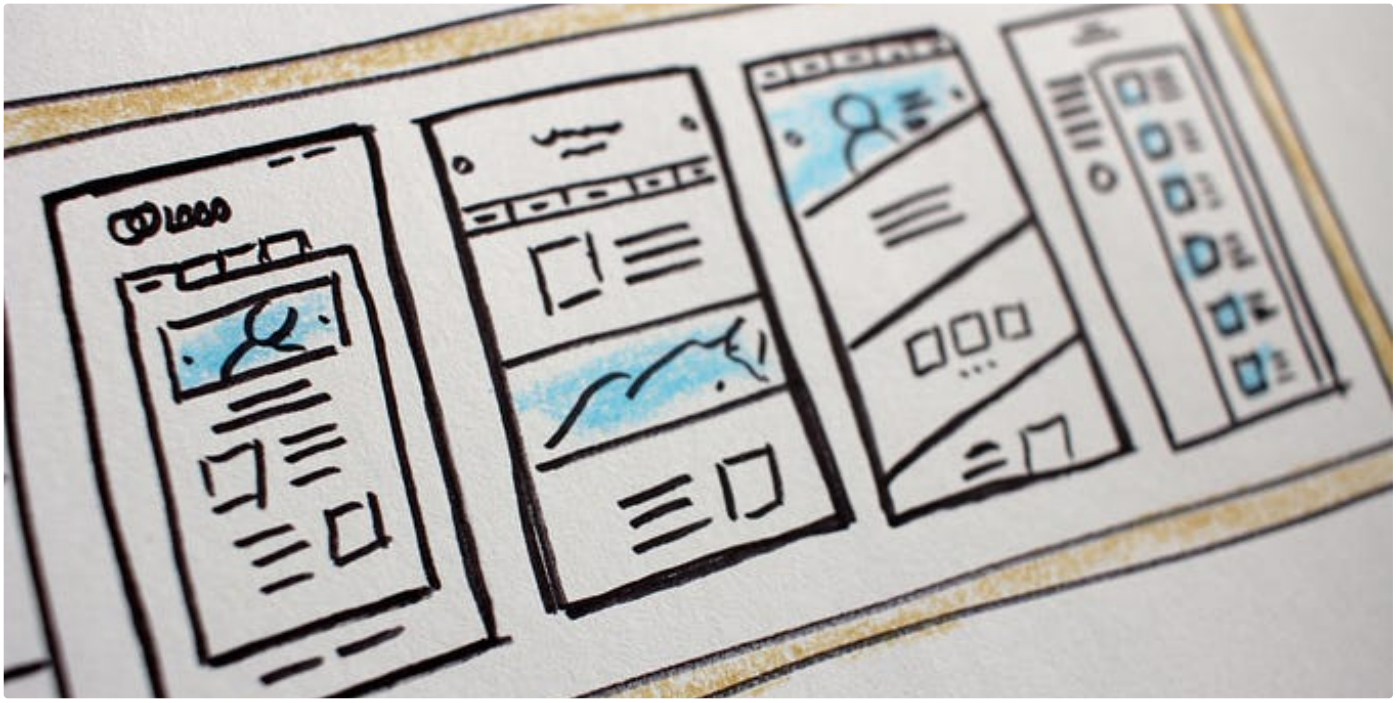
Bernardo Iribarne in Nerd For Tech

Flutter error handling: a cleaned design to manage errors in your apps using the Decorator design...

Handling errors like a senior developer

7 min read · Nov 2

111



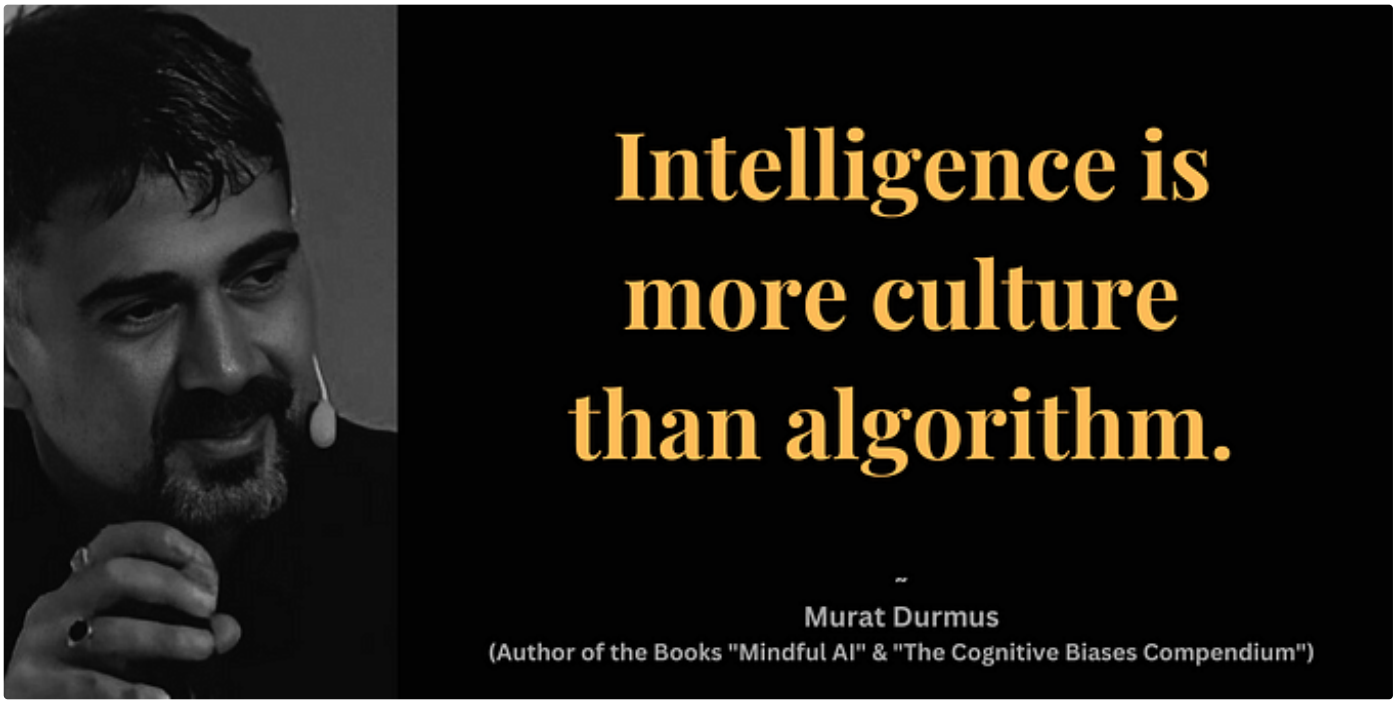
Bernardo Iribarne in Nerd For Tech

Layouts in Flutter: how to implement page layouts using Template Method design pattern

Building page layouts in Flutter

5 min read · Nov 15

81 2



Murat Durmus (CEO @AISOMA_AG)

Intelligence is more Culture than Algorithm

Intelligence is not just a question of pure data processing or logical algorithms. Instead, it encompasses a broader range of abilities...

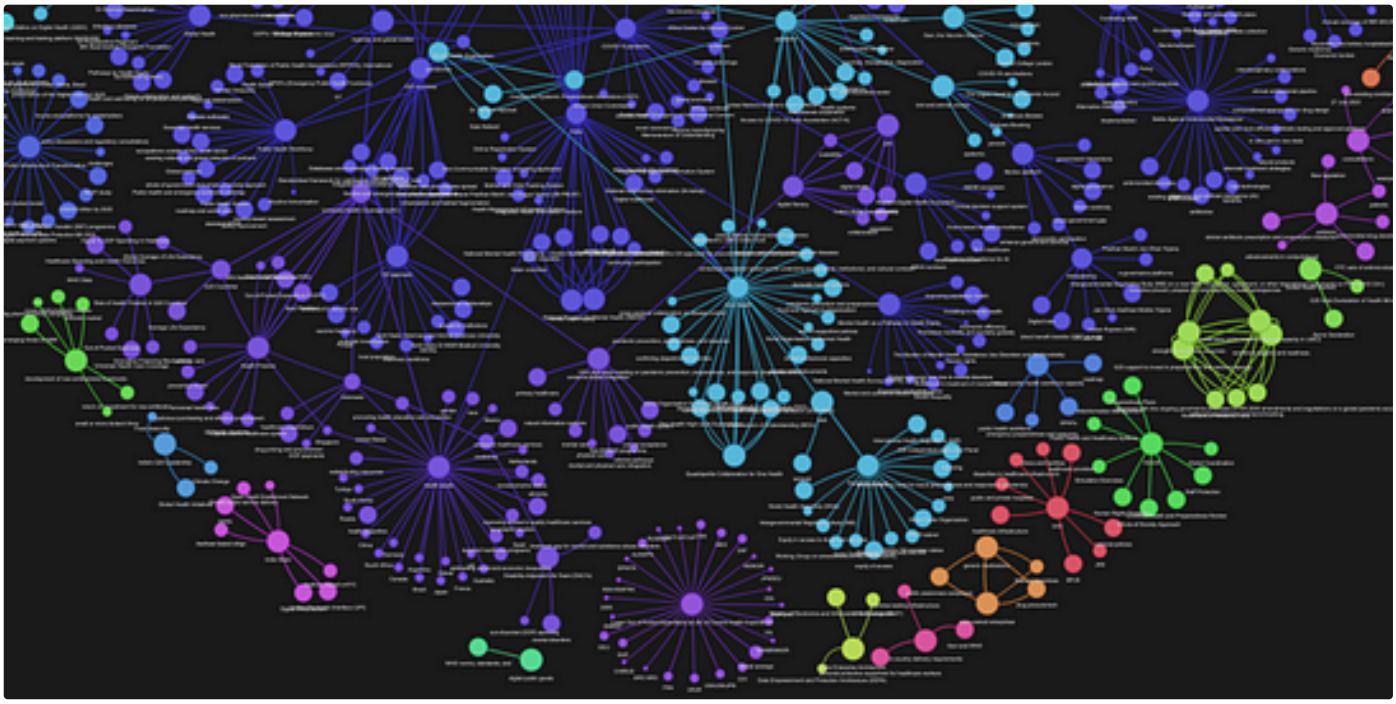
1 min read · Nov 8

16

See all from Murat Durmus (CEO @AISOMA_AG)

See all from Nerd For Tech

Recommended from Medium



Rahul Nayak in Towards Data Science

How to Convert Any Text Into a Graph of Concepts

A method to convert any text corpus into a Knowledge Graph using Mistral 7B.

12 min read · Nov 10

3.3K 37

Thomas Smith in The Generator

Did OpenAI Secretly Create a Brain-Like Intelligence After All?

What We Know About Q* and OpenAI's Potential AGI Breakthrough

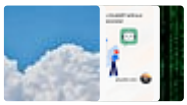
· 5 min read · 5 days ago

2.9K 48

Lists

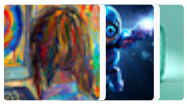
AI Regulation

6 stories · 206 saves



ChatGPT

22 stories · 286 saves

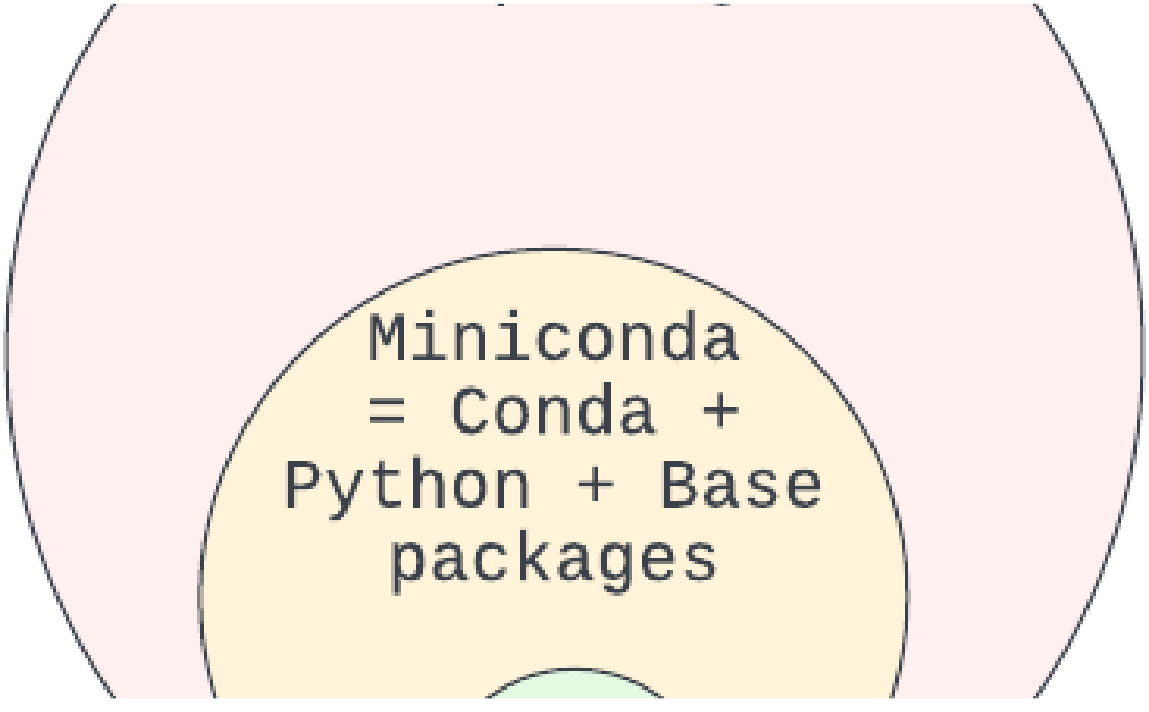


ChatGPT prompts

30 stories · 737 saves

Generative AI Recommended Reading

52 stories · 459 saves



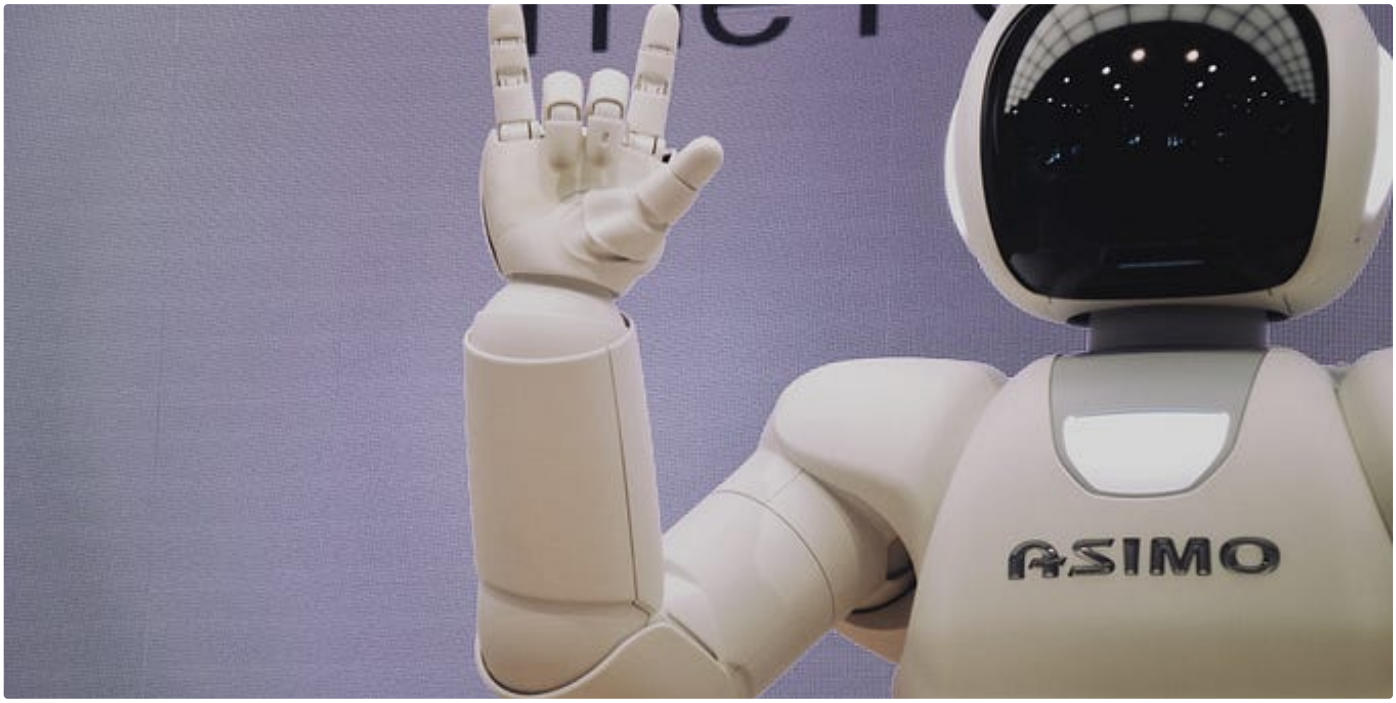
Chaitanya (Chey) Penmetsa in CodeNx

Environment setup and troubleshooting for Machine Learning

Having gained an understanding of Machine Learning and established the framework for ML projects, let's shift our attention to the tools...

9 min read · Nov 23

50

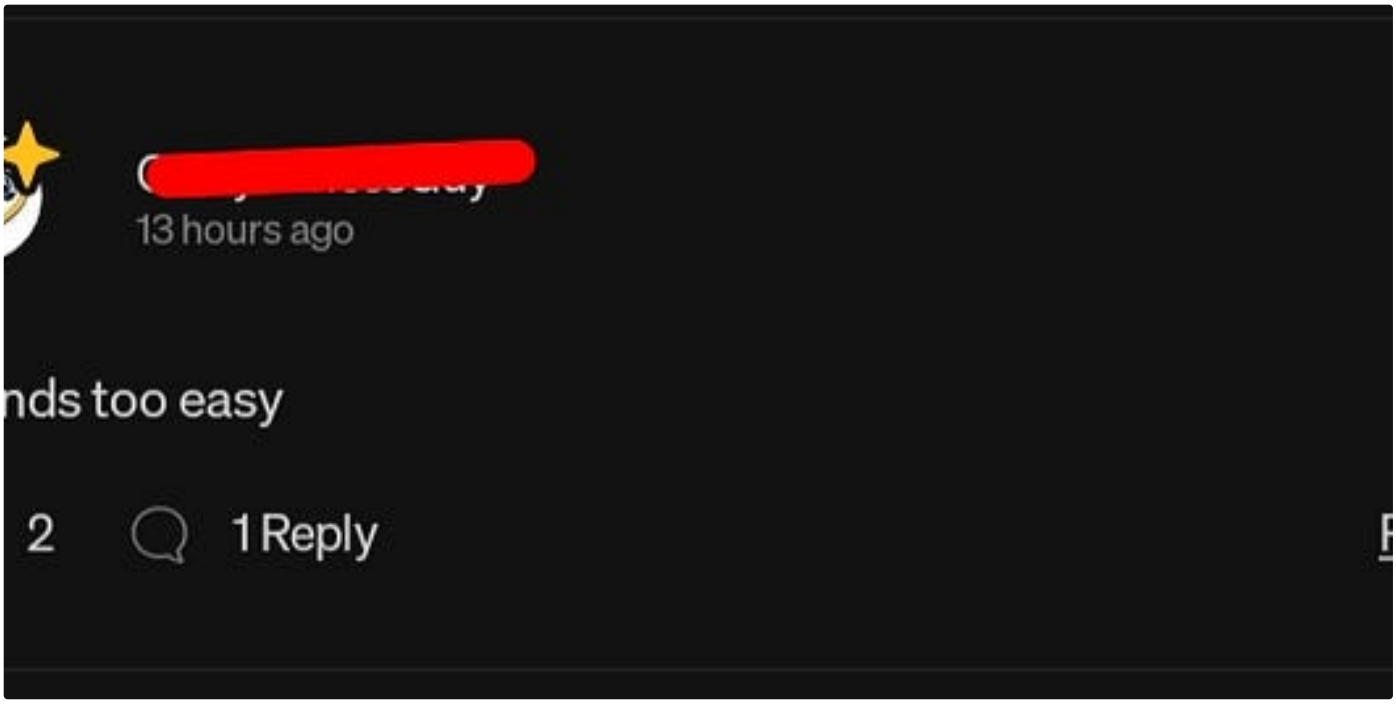


TheTechPencil in *AI monks.io*

The Potential Risks and Benefits of Artificial General Intelligence (AGI)

Navigating the Future of AGI: A Balanced Approach to Harnessing Power and Mitigating Risks

· 4 min read · Jun 21



Paul Rose

I Found A Very Profitable AI Side Hustle

And it's perfect for beginners

6 min read · Oct 18

12.2K 216

Oops!

Access is temporarily unavailable, please try again later.

Please contact us through our help center if this issue persists.

Jeremy Arancio in Towards AI

Why are AI Products Doomed to Fail?

After one year of implementing AI features for various businesses, I share my perspective on the mistakes I see companies making with LLMs...

✨ · 17 min read · Nov 17

 1.1K  28 

See more recommendations