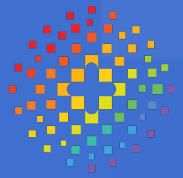


edge ai + vision
ALLIANCE™

Your Next Computer Vision Model Might be an LLM: Generative AI and the Move From Large Language Models to Vision Language Models

October 23, 2024

Welcome!

Jeff Bier

Edge AI and Vision Alliance

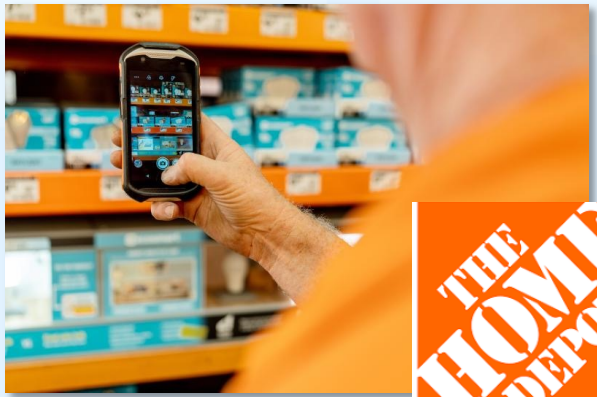
Solving Real-World Problems at Scale

Philadelphia Enquirer

Verily

Wabtec

BNSF.com



Chain Store Age

© 2024 Edge AI and Vision Alliance

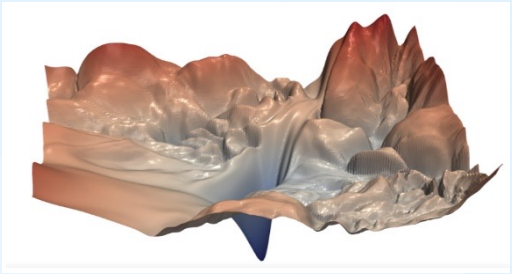
Inspire and Empower Product Creators

- Learn about new technologies, techniques and capabilities
- Build skills and know-how
- Connect with colleagues, suppliers, open source projects, standards and other resources

© 2024 Edge AI and Vision Alliance

Resources for Product Creators

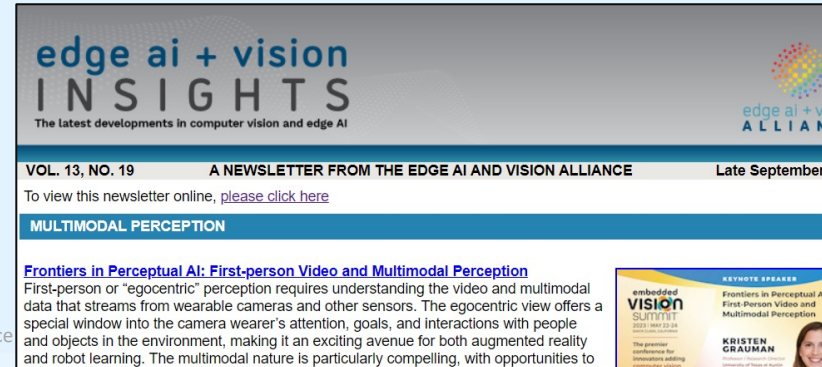
Navigating the Future: How Avnet is Addressing Challenges in AMR Design



Quantization of Convolutional Neural Networks: Quantization Analysis

Edge AI and Vision Alliance Conversation with GenAI Nerds on Generative AI At the Edge

www.edge-ai-vision.com


edge ai + vision
INSIGHTS
The latest developments in computer vision and edge AI

VOL. 13, NO. 19 A NEWSLETTER FROM THE EDGE AI AND VISION ALLIANCE Late September

To view this newsletter online, please click here

MULTIMODAL PERCEPTION

Frontiers in Perceptual AI: First-person Video and Multimodal Perception
First-person or "egocentric" perception requires understanding the video and multimodal data that streams from wearable cameras and other sensors. The egocentric view offers a special window into the camera wearer's attention, goals, and interactions with people and objects in the environment, making it an exciting avenue for both augmented reality and robot learning. The multimodal nature is particularly compelling, with opportunities to

© 2024 Edge AI and Vision Alliance

Helping Companies Grow

- Gain insights into trends in markets, technologies, applications and standards
- Connect with customers, suppliers and ecosystem partners
- Become visible as a thought leader

© 2024 Edge AI and Vision Alliance

Edge AI and Vision Alliance Member Companies

© 2024 Edge AI and Vision Alliance

edge ai + vision ALLIANCE™

Inspiring + empowering innovators to design systems that **perceive + understand**

Visit *membership*.edge-ai-vision.com

© 2024 Edge AI and Vision Alliance

May 20-22

The premier conference for innovators incorporating computer vision and AI in products.

embeddedvisionsummit.com

Brought to you by the

edge ai + vision ALLIANCE™

Agenda

Today's Objectives

- Introduce vision language models (VLMs) and large multimodal models (LMMs), and how they relate to LLMs
- Explain why and how VLMs and LMMs are becoming important for computer vision
- Illustrate real-world VLM and LMM uses cases
- Introduce how VLMs and LMMs work and how they can be incorporated into applications
- Identify challenges in using VLMs and LMMs
- Answer your questions

© 2024 Edge AI and Vision Alliance

Today's Agenda

- **Introduction**: Jeff Bier, Edge AI and Vision Alliance
- **Real-world LMM use cases**: Carter Maslan, Camio
- **How LMMs work and how to use them** (Part 1!): István Fehérvári, BenchSci
- **Discussion and Q&A**

Jeff Bier
Edge AI and Vision Alliance

Carter Maslan
Camio

István Fehérvári
BenchSci

© 2024 Edge AI and Vision Alliance

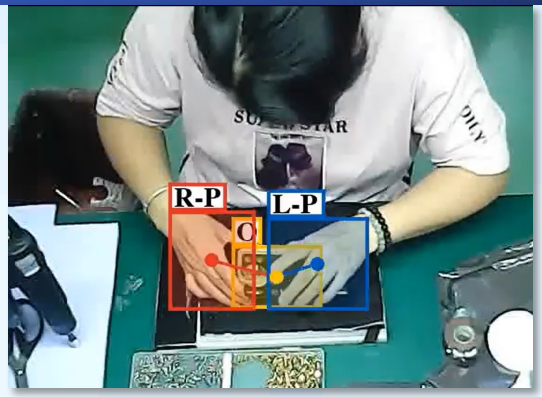
From CNNs to LLMs to LMMs

edge ai + vision ALLIANCE™

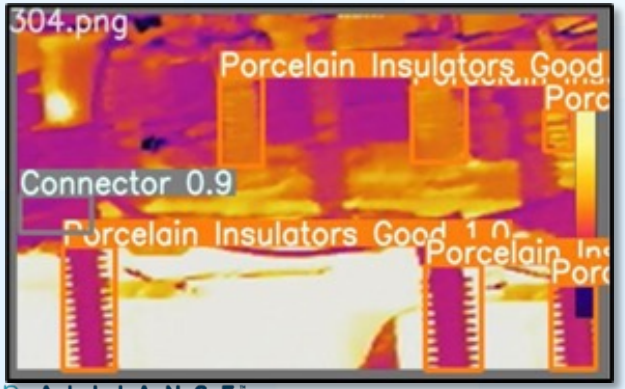
Inspiring + empowering innovators to design systems that **perceive + understand**

Deep Learning Enables Computer Vision to Work In the Real World

Dexterity, Inc.



Retrocausal

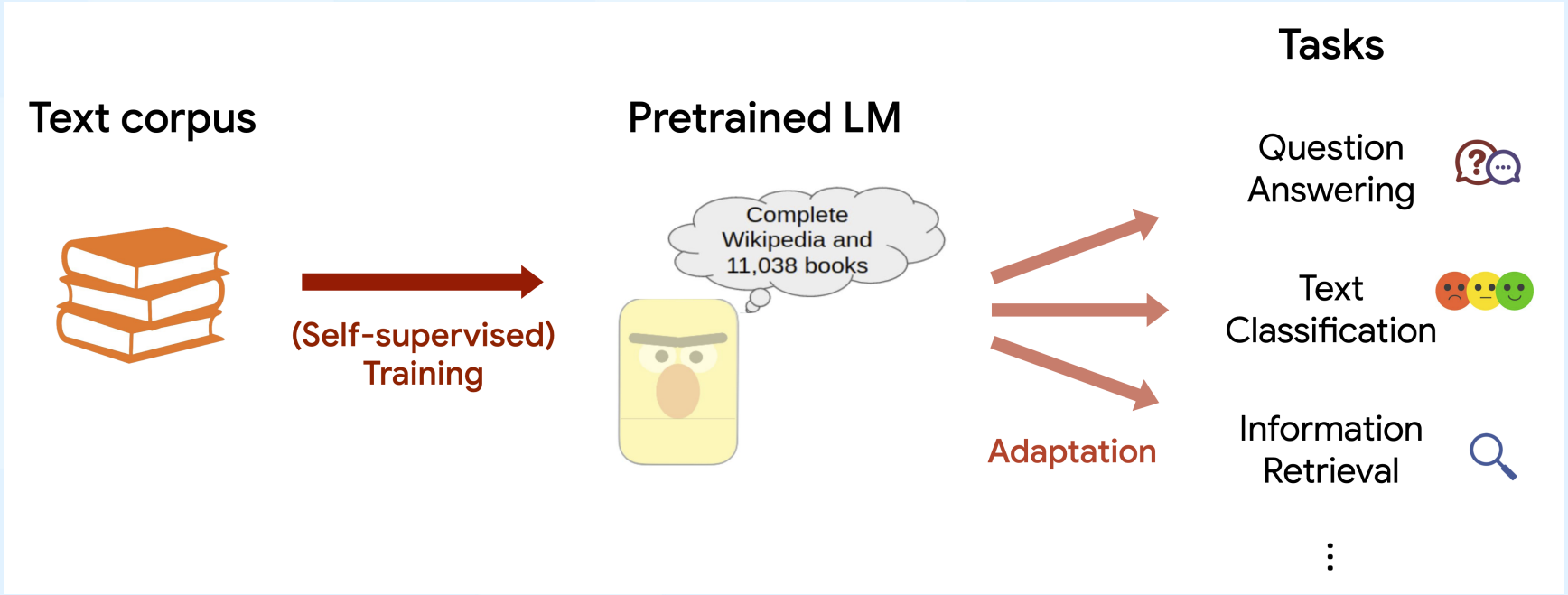


Buzz Solutions

Blue River Technology

© 2024 Edge AI and Vision Alliance

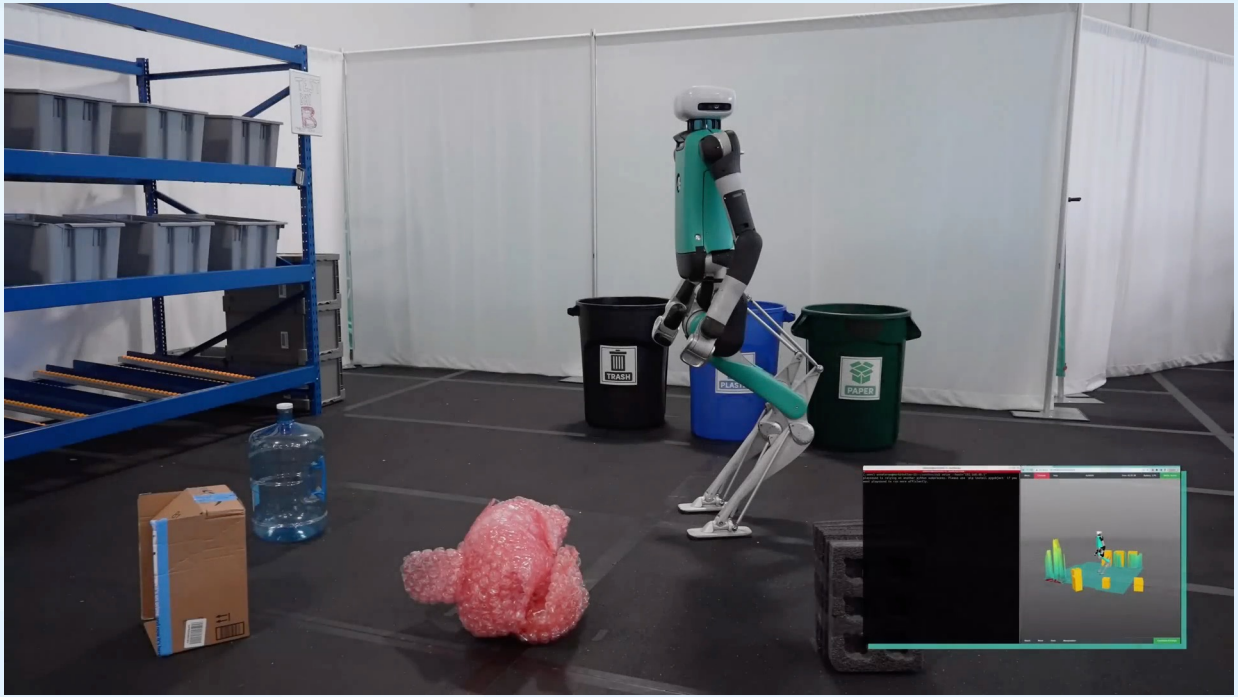
Large Language Models



Everton Gomede, PhD

© 2024 Edge AI and Vision Alliance

How Will Generative AI Change Perceptual AI?

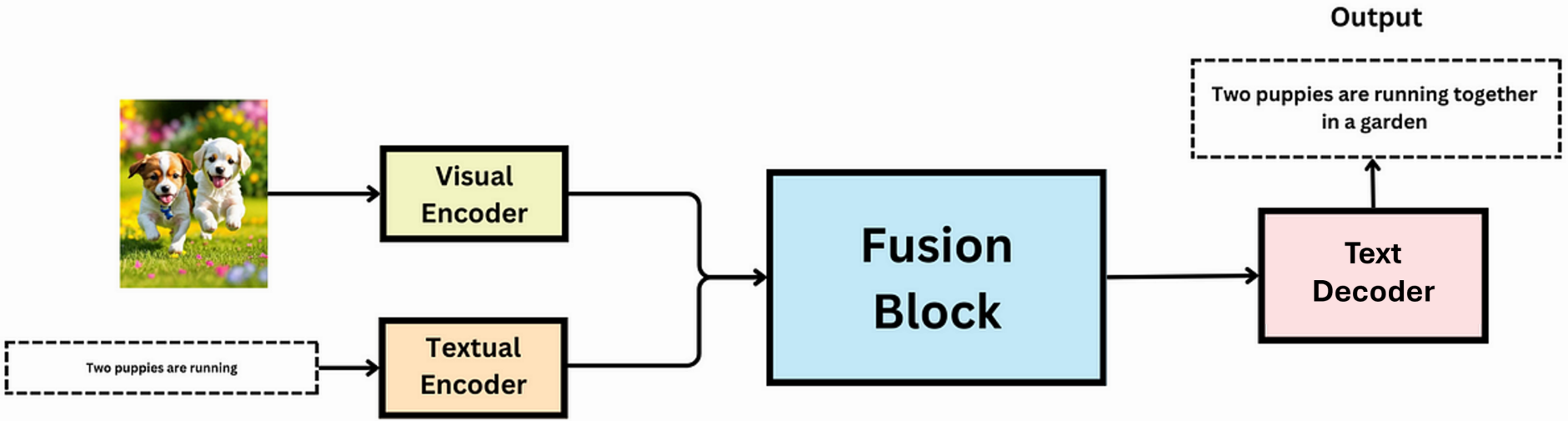


Source: Agility Robotics

https://youtu.be/Vq_DcZ_xc_E

© 2024 Edge AI and Vision Alliance

Vision Language Model Architecture



Original figure by Prashant Kalepu, Medium – modified by Jeff Bier

© 2024 Edge AI and Vision Alliance

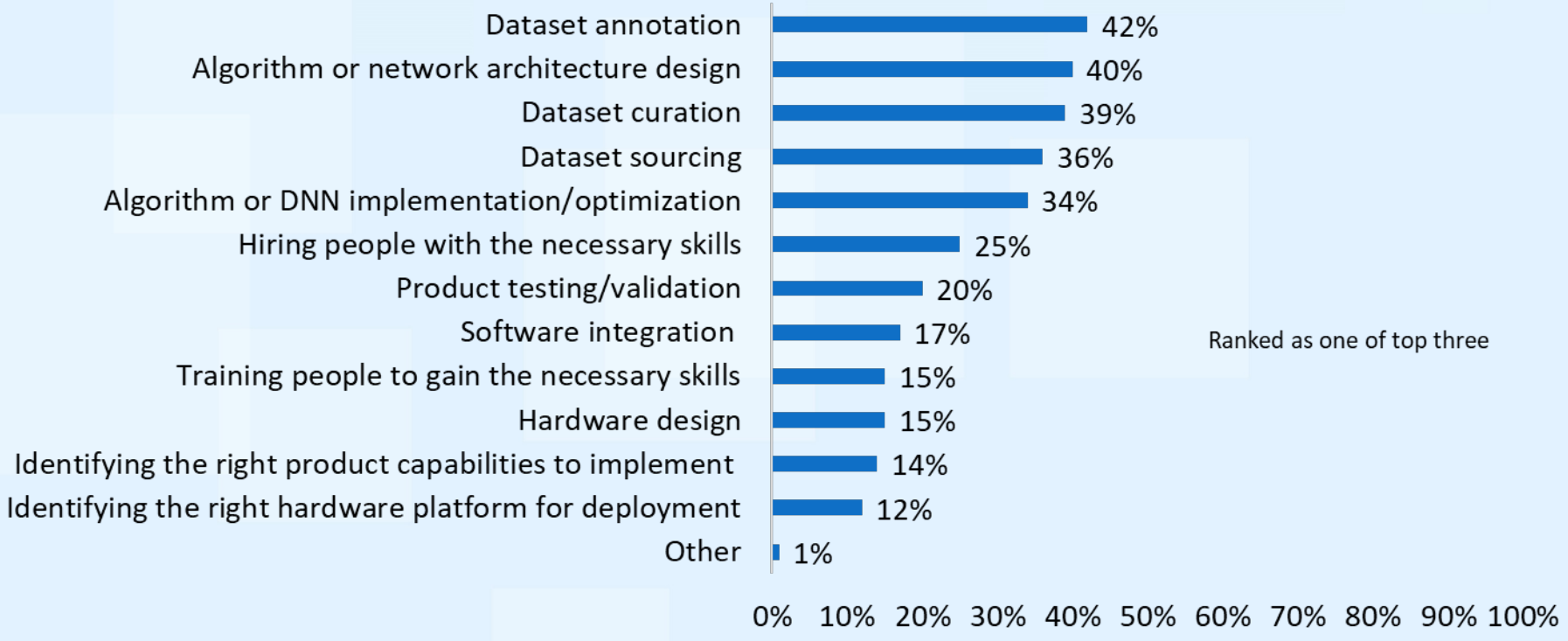
VLM and LLM Advantages

- Off-the-shelf foundation models can be used for many applications
- Query image/video data via language
- Get language descriptions of images/video
- Multimodal perception leads to better perception and understanding
- Can be better at generalizing
- Can be better at distinguishing subtle differences, including in behavior of people
- Enable application developers to work at higher levels of abstraction
- Can act as controllers, calling other models (e.g., for counting, pose) and deciding on appropriate actions

© 2024 Edge AI and Vision Alliance

The Training Data Challenge

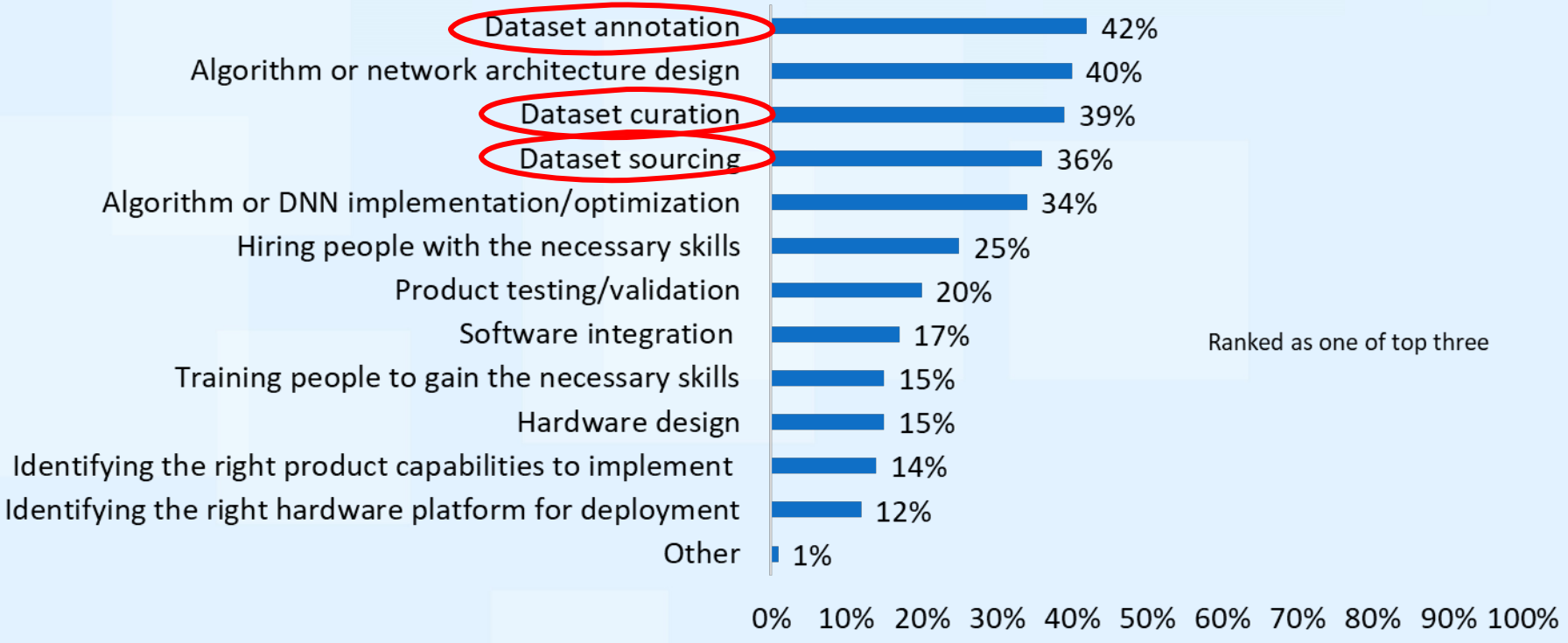
Areas of Computer Vision/Machine Perception Product Development Most Challenging



Source: Edge AI and Vision Alliance, *Computer Vision and Perceptual AI Developer Survey, November 2023*

© 2024 Edge AI and Vision Alliance

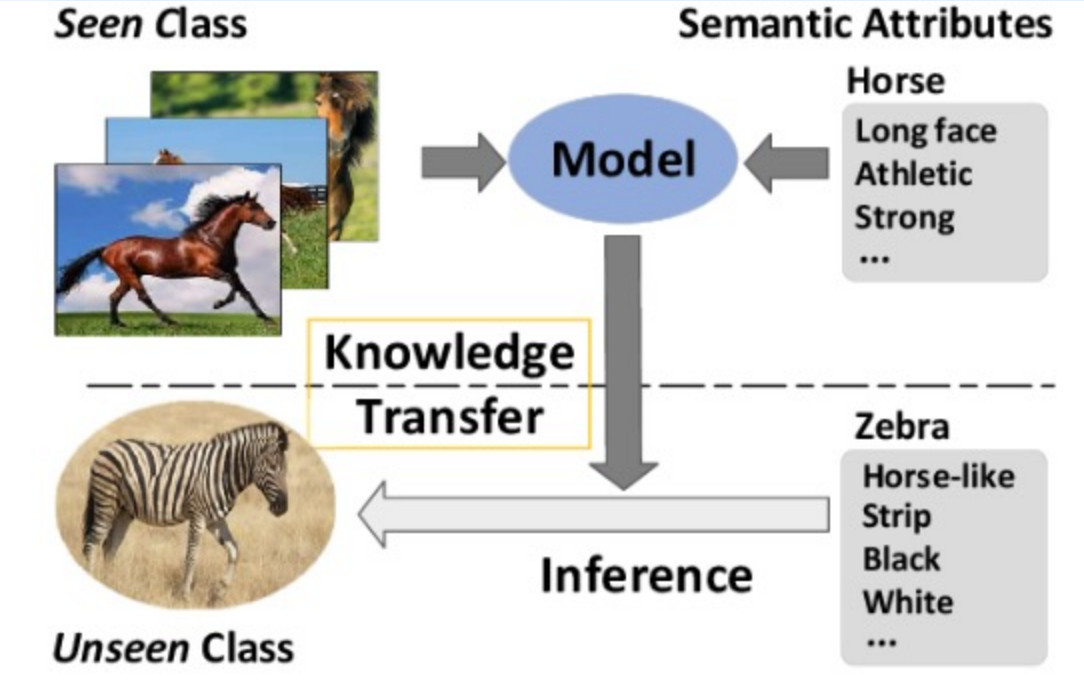
Areas of Computer Vision/Machine Perception Product Development Most Challenging



Source: Edge AI and Vision Alliance, *Computer Vision and Perceptual AI Developer Survey, November 2023*

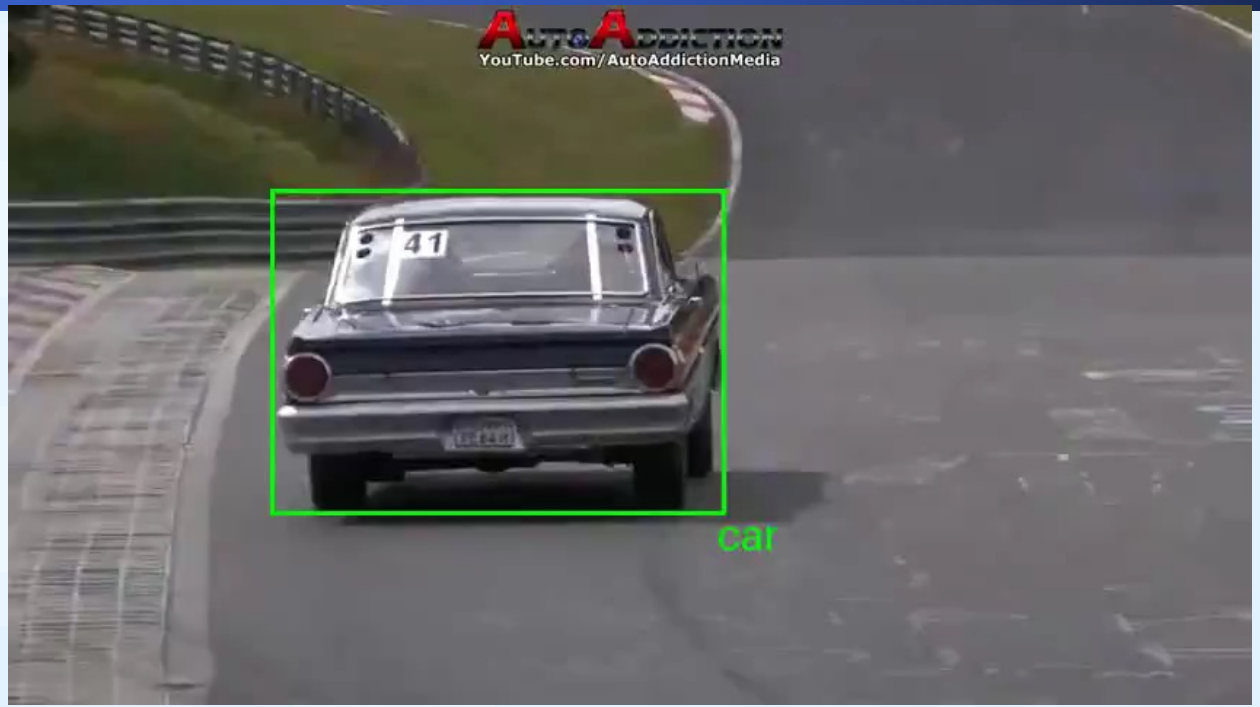
© 2024 Edge AI and Vision Alliance

Zero-Shot Learning



Source: Sarojag, Analytics Vidhya

© 2024 Edge AI and Vision Alliance



@sumo43

x.com/sumo43_/status/1791589684121903555

© 2024 Edge AI and Vision Alliance

Generalization Is Key for Many Real-World Vision Applications

How LLMs Are Changing Computer Vision

Visual input example, Extreme Ironing:

Source: https://www.barnorama.com/wp-content/uploads/2016/12/03-Confusing-Pictures.jpg

LLaVA

What is unusual about this image?

The unusual aspect of this image is a man ironing clothes on the back of a minivan or van. This is not a typical place to perform this activity, as one would usually iron clothes in a more stationary and safe location, such as a home, using a regular ironing board. The scene depicted in the image is peculiar as it involves a makeshift ironing setup on a vehicle, which can be both unsafe and unconventional. Additionally, it is not clear how the man is able to maintain balance and stability while ironing clothes in such an unstable environment.

Yong Jae Lee

© 2024 Edge AI and Vision Alliance

How LLMs Are Changing Computer Vision

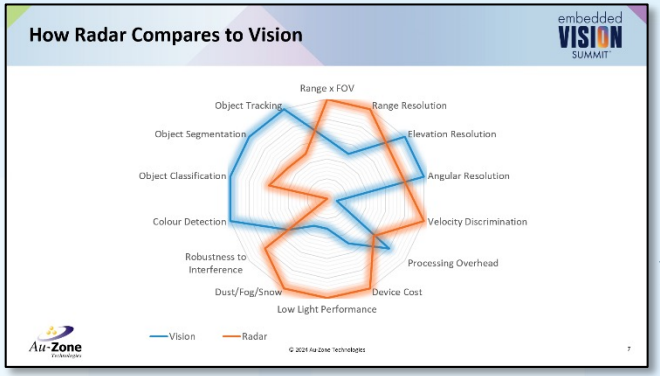
Vertikal.net

© 2024 Edge AI and Vision Alliance

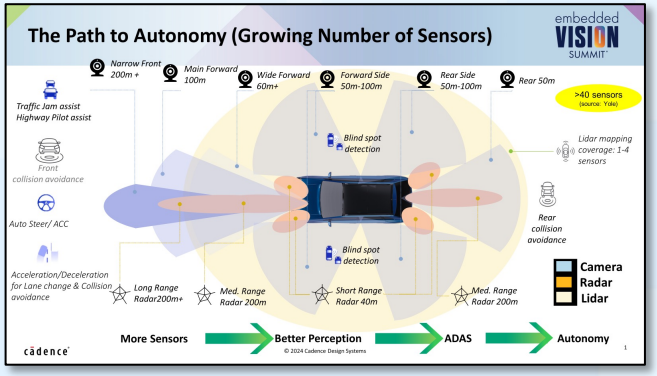
The Need for Multimodal Perception

Challenges – The Need for Multimodal Perception

All About Vision



Sebastien Taylor, Au-Zone



Amol Borkar, Cadence

Waymo

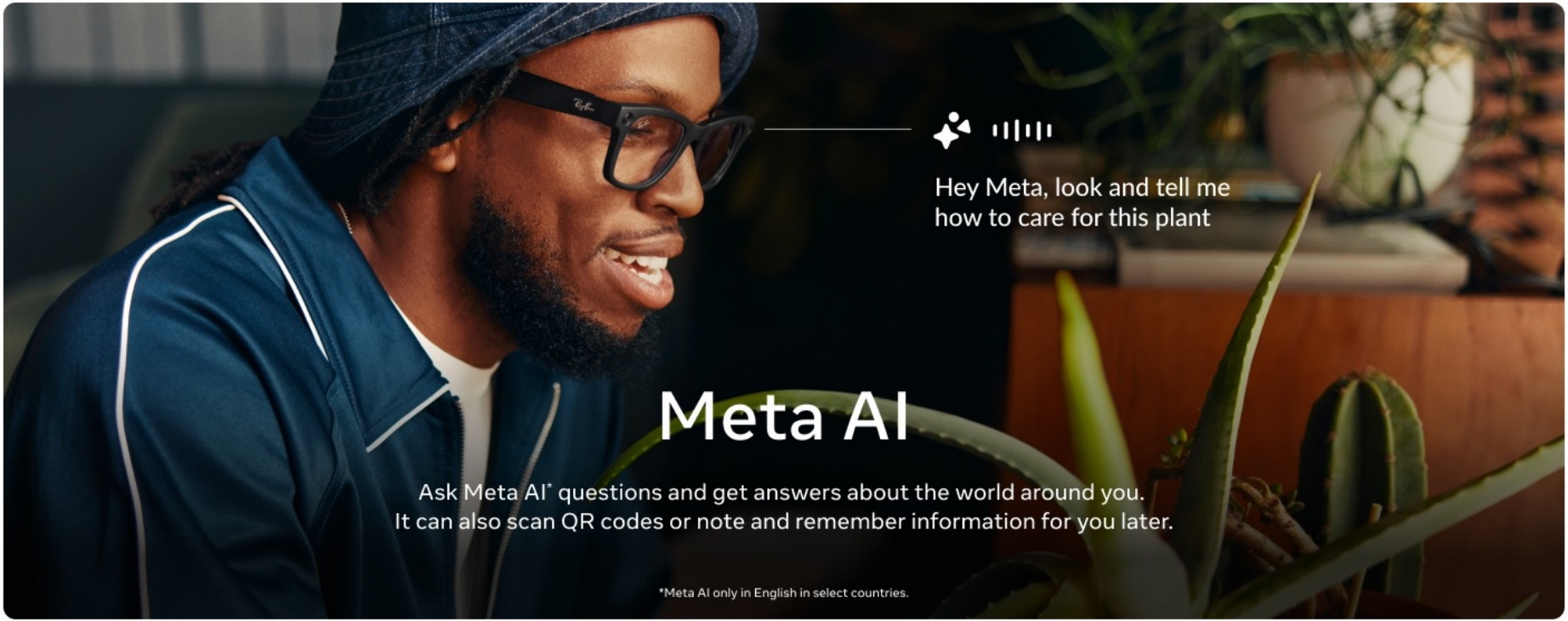
"With 13 cameras, 4 lidars, 6 radars, and an array of external audio receivers, our [6th generation] sensor suite is optimized for greater performance at a significantly reduced cost, without compromising safety."

-Waymo

© 2024 Edge AI and Vision Alliance

LMMs in Commercial Applications

Ray Ban Meta Smart Glasses



Demo video at www.youtube.com/watch?v=NOdjgw_v4bw

© 2024 Edge AI and Vision Alliance

Challenges

Challenges Incorporating VLMs and LLMs into Products

- It can be difficult to select the best model for an application
- Models may require customization for a given application
- Most developers aren't familiar with how to select, customize and use these models
- Models are very large (memory, memory bandwidth, compute)
  - Deployment at the edge is bleeding-edge; running in the cloud is costly
- General-purpose nature does not mean they can do everything, or do everything well
  - Often ill-suited to specific tasks, such as object counting or pose estimation
- Many widely used models are closed

© 2024 Edge AI and Vision Alliance

Your Next Computer Vision Model Might be an LLM

Generative AI and the Move From Large Language Models to Vision Language Models

Jeff Bier
Founder and President
Edge AI and Vision Alliance

Carter Maslan
CEO
Camio

István Fehérvári
Director of Data and ML
BenchSci

© 2024 Edge AI and Vision Alliance

Discussion and Q&A

edge ai + vision ALLIANCE™

Inspiring + empowering innovators to design systems that **perceive + understand**

Thank You!

edge ai + vision ALLIANCE™

Inspiring + empowering innovators to design systems that **perceive + understand**

Visit edge-ai-vision.com

© 2024 Edge AI and Vision Alliance