Making Alexa More Ambiently Intelligence with Computer Vision

Michael Giannangeli

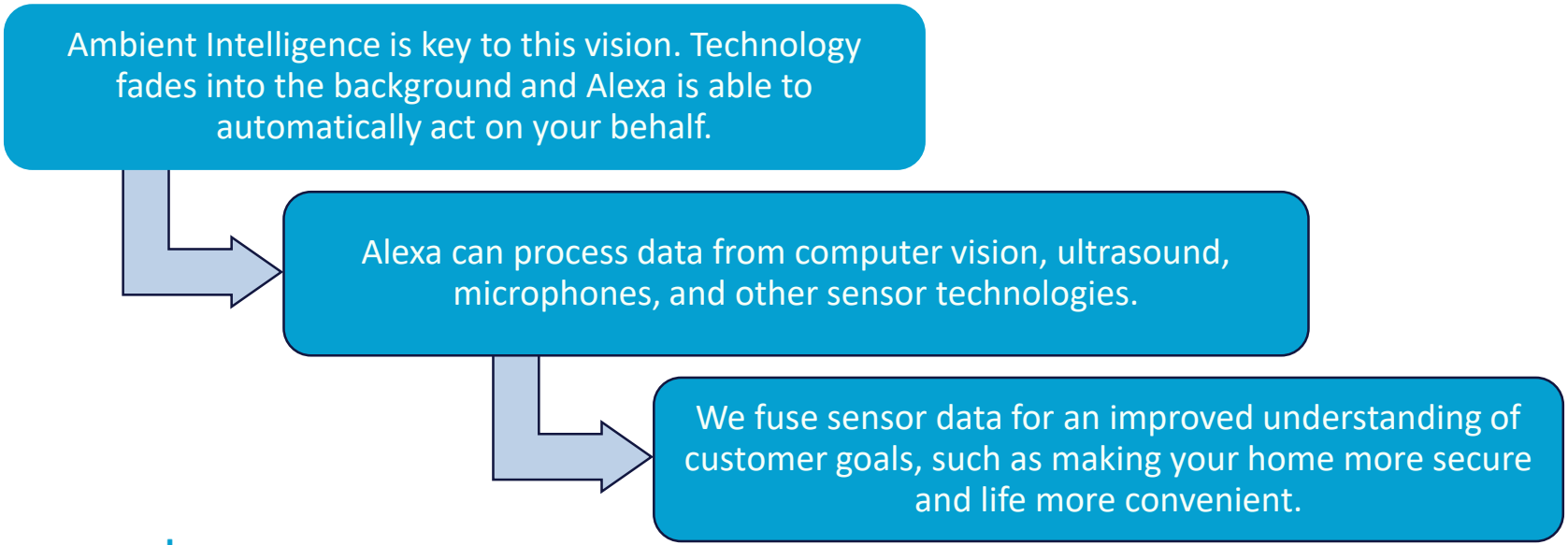
Senior Manager, Product Management, Alexa Devices

Amazon Lab126

amazon alexa

Alexa's Vision

Alexa is your trusted assistant, advisor, and companion that is always finding ways to make life more convenient and fulfilling.



Innovative Features Using Computer Vision

- **Adaptive Content**: Uses CV to detect proximity and dynamically change the screen content. When a customer is nearby, more details and touch targets are surfaced; when far away, text is larger to make it easier to see.
- **Smart Motion**: Fuses microphone and CV input to move with you and keep you in the Echo Show 10’s field of view during video calls or while you're cooking along to a recipe.
- **Visual ID**: Opt-in feature designed to recognize you and your family members so Alexa can show personalized content such as calendars and reminders, recently played music, news, and notes for you.

Case Study: Adaptive Content

Working Backwards from the Customer

- All starts with a PRFAQ - press release and frequently asked questions.
- Forcing function to work backwards from the customer. Aligns leadership on a vision.
- Not everything in this document came true (including the target launch date) and throughout development the team must re-align as new data emerges.
- As part of this process, the product manager defines requirements, KPIs, and success metrics.
  - Accuracy: Precision/Recall
  - Detection Range
  - User Perceived Latency
  - Model Resources (RAM/CPU)

CUSTOMER PRESS RELEASE

Amazon introduces Adaptive UI on Echo Show to ambiently change content based on user proximity

Alexa ambiently changes the content on the screen as you get closer to the device, exposing touch controls and showing more details

When you're further from the device, Alexa makes it easier to see content like your timers and smart home devices

SEATTLE – (BUSINESS WIRE) – October 1, 2022 — Amazon [NASDAQ: AMZN] today announced Adaptive UI, a new feature on Echo Show devices that ambiently changes the content on the screen based on your proximity to the device. When you're nearby, Adaptive UI invites you to interact with your favorite content and ensures you don't miss timely information. For example, when media is playing and a customer is within 4 feet of the device, touch controls automatically surface and pending notifications appear. On the other hand, when you're further from the device, Adaptive UI makes important content larger so it is easier to see, such as your timers counting down. This is especially helpful on the new Echo Show 5 (3rd Gen). "I was interested in buying the new Echo Show 5, but was afraid that the screen would be too small for me to see the display and clock from a distance, especially with my poor eyesight," said John Overturf from Arizona. "When I saw it featured Adaptive UI, I immediately purchased it."

Easy to read from a distance: When you're further away from your device, Adaptive UI increases the size of the Home screen content, such as the time and weather, to make it easier to read from a distance.

Music touch controls: When playing music through your Echo Show on the Now Playing screen, touch controls automatically surface when you're near the device. Touch controls to pause, skip, or go back surface when music is playing and controls to resume surface when music was recently paused.

Adaptive Smart Home controls: When you open the Smart Home dashboard and you're near your Echo Show, more of your compatible devices such as lights, cameras, thermostats, and plugs are shown allowing you to easily toggle them on or off. When you're further away, fewer devices are shown larger, such as your favorites, making it easier to see and interact with by voice. Your Smart Home widget is also adaptive, showing a large read-only version from a distance and surfacing touch controls as you get closer.

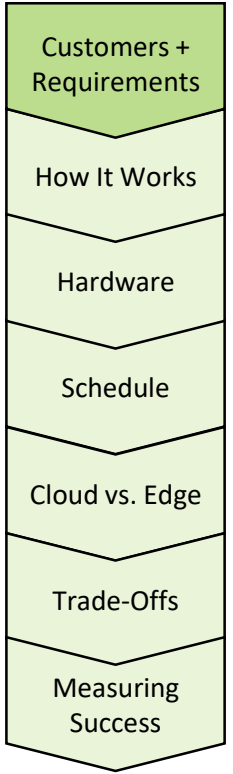
Enhanced shopping experience: When you shop with Alexa, different UI is shown depending on if you're near or far from the device. If you're near, multiple options for purchase will be displayed, including the previously ordered ASIN, a different brand, different pack size, or similar items that are on sale. Touch controls are also surfaced to make it easier to select a different option or keep browsing. If you're further away, only one purchase option will be shown with a large image and font size, making it easier to read and quickly order by voice. "I was worried that with Adaptive UI, things would move around and throw me off as I walked towards the screen. But the transitions are so smooth and seamless that I don't lose what I was looking at," said Chrissy Ross.

Quickly view your notifications: Pending notifications will appear when you're near the device, ensuring you won't miss urgent information from Alexa. If you're also enrolled in visual ID, only your notifications will show up when you're recognized.

Better control of your timers: With Adaptive UI, your timer is shown more prominently when you're further away from the device and controls to dismiss the timer are surfaced when you're nearby. If you have multiple timers, all timers are shown together. "I like that Amazon introduced the persistent timer pill last year, but usually I have multiple timers running while cooking and it becomes hard to see them when I'm more than a few feet from the device," said Marjorie Wilson, a mother of two. "Now Alexa makes it so much easier to see my timers from a distance and interact with them when I'm closer, like deleting the one incorrect timer I accidentally set."

Intuitive feature discovery: Adaptive UI allows customers to discover new functionality by exposing certain touch controls that they may not have known existed, such as media transport controls or the ability to share recipes.

"Adaptive UI furthers our vision of ambient intelligence for Alexa," said Tom Taylor, Senior Vice President of Alexa. "Our goal is that customers won't even realize that Adaptive UI is enabled - the experience is so seamless and effortless - and is yet another example of Alexa getting smarter to make your experience better." Adaptive UI is available on the new Echo Show 5 (3rd Gen), as well as Echo Show 15, Echo Show 10 (3rd Gen), and Echo Show 8 (2nd Gen). For more information, visit Amazon.com/AdaptiveUI.



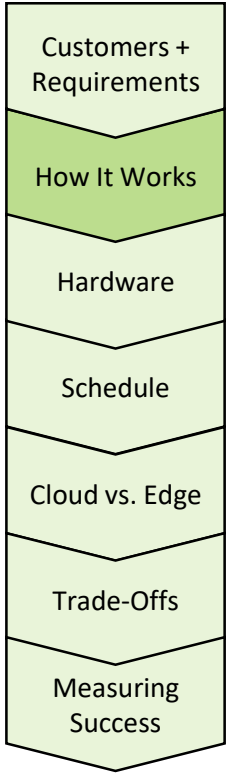
How Does Adaptive Content Work?

- Adaptive Content uses a three-stage computer vision model:
    1. Stage one looks for a person
    2. Stage two looks for a head
    3. Stage three uses the size of the head to infer distance to the device
- Our CV service turns the camera image into hundreds of data points representing shapes, edges, facial landmarks, and general coloring.
- All processing happens on device in milliseconds. Then the image is deleted permanently.

- Customers + Requirements
- How It Works
- Hardware
- Schedule
- Cloud vs. Edge
- Trade-Offs
- Measuring Success

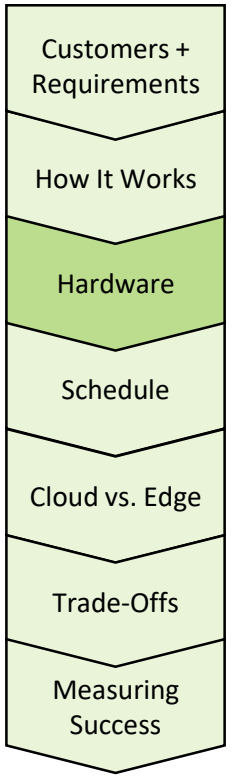
Computer Vision Challenges

- When multiple people are in the field-of-view, who do you prioritize?
  - The closer person or the person actively engaged with the device?
  - What if there is someone nearby who touches the screen but isn't in the field-of-view?
- Adults and kids will have different experiences.
  - Inferring distance based on size of the head is an imperfect science. What is considered the average person and does that result in an equitable outcome for diverse customers?
- Won't work in a very dark room or when the camera shutter is closed.
  - Does touch become the signal to change to near-field? Or do we turn the feature off?



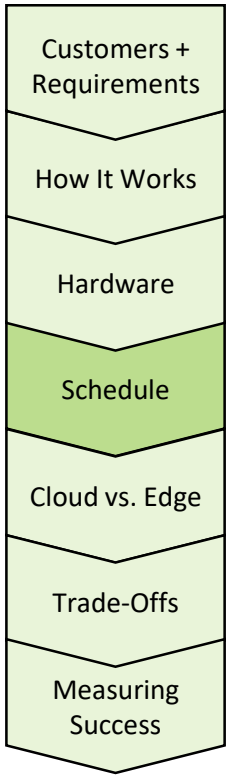
Hardware Constraints

- Can you influence the hardware or are you launching your feature on an existing device?
- Consider what hardware capabilities are available?
  - Memory / Compute
  - Camera (megapixels, low-light performance)
- If you want to add hardware capability to support your feature, you need to justify the cost.
  - Does your feature lead to more sales? A higher sales price? Increased engagement? New streams of monetization?



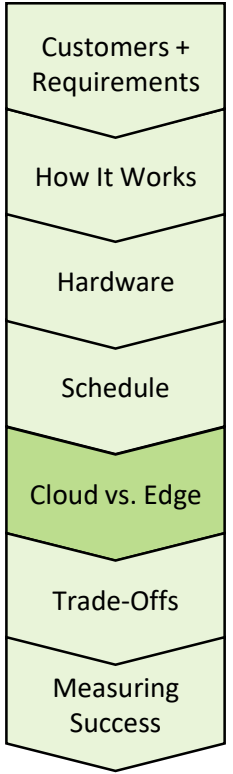
Managing Schedule

- When do you need to launch your feature? Is it tied to a device launch?
- Create a working backwards schedule
- Long-poles in development:
  - Data Collection
  - Model Development
  - Device Integration
  - Alpha/Beta Testing



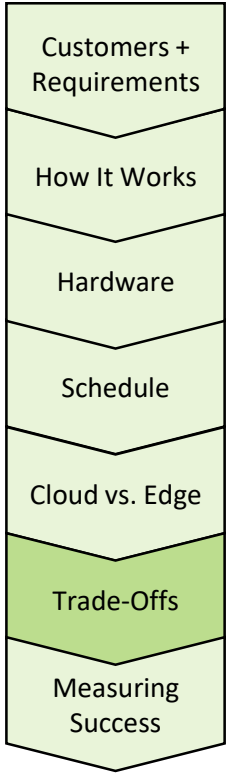
Cloud vs. Edge Processing

- How much memory headroom is available on device? Can you store and run the CV model locally?
- How important is latency? Can you afford the increased latency that comes from running a model in the cloud?
- Is this a feature in which customers value privacy? Will you lose customers if their images are sent to the cloud?



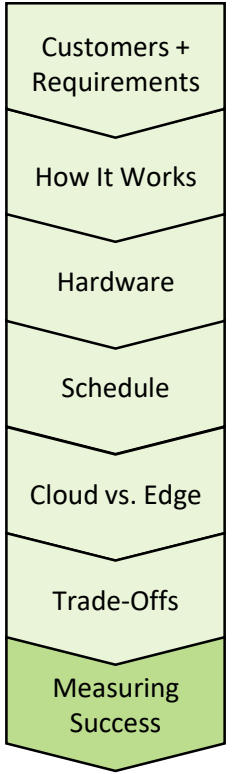
Trade Offs

- It's the Product Manager's job to assess these different vectors and make the best decision for customers and the business.
- Performance vs. cost vs. schedule vs. privacy:
  - Do you add cost for an improved camera or low-light sensing?
  - Do you add memory so the device can run more models concurrently and have faster latency? Do you sacrifice model accuracy for a smaller model?
  - Can you reduce pre-launch development time by continuing to refine and improve your model post-launch with additional data collection?
  - Are privacy and latency worth increasing the compute and memory?



Measuring Success

- High Engagement: We compared engagement on Echo Show 8 (3rd Gen) and Echo Show 8 (2nd Gen), controlling for “newness”, and saw more than a **25% increase in actions per user**.
- High Adoption: Adaptive Content is a default-on feature. **Less than 1% of customers have turned the feature off (and most of them have turned it back on)**, providing a strong signal that customers find it valuable.



Measuring Success

- Tech media has also had very positive reviews:
  - ZDNET wrote “Adaptive Content will make it easier for consumers to view content on the screen from a distance by simplifying it when no one is near the device and switching to a detailed view when the person approaches the Echo Show.”
  - PCMag called Adaptive Content a “nice touch”.
  - TheStreet called it “the neatest part of the display”.
  - Android Headlines said it was “one of those features that just makes so much sense, and it’s one that you don’t even think about.”