Assessing Dialog Management Systems

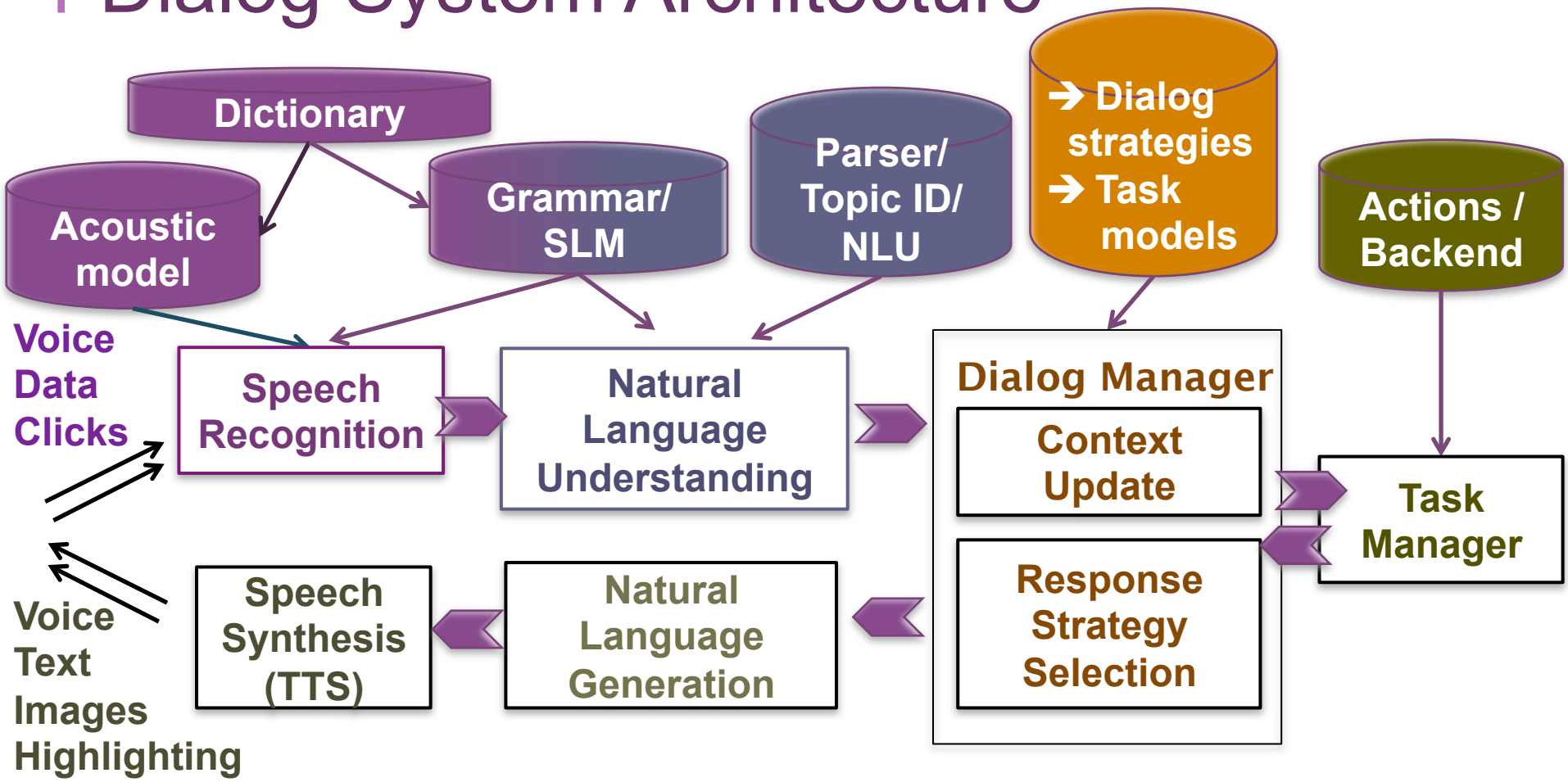
Presenters:
K.W. (Bill) Scholz
Marie Meteer

With help from the rest of the
AVIOS Advanced Dialog Group:
Emmett Coin
John Tadlock
Lorin Wilde

From ASR to Dialog Management

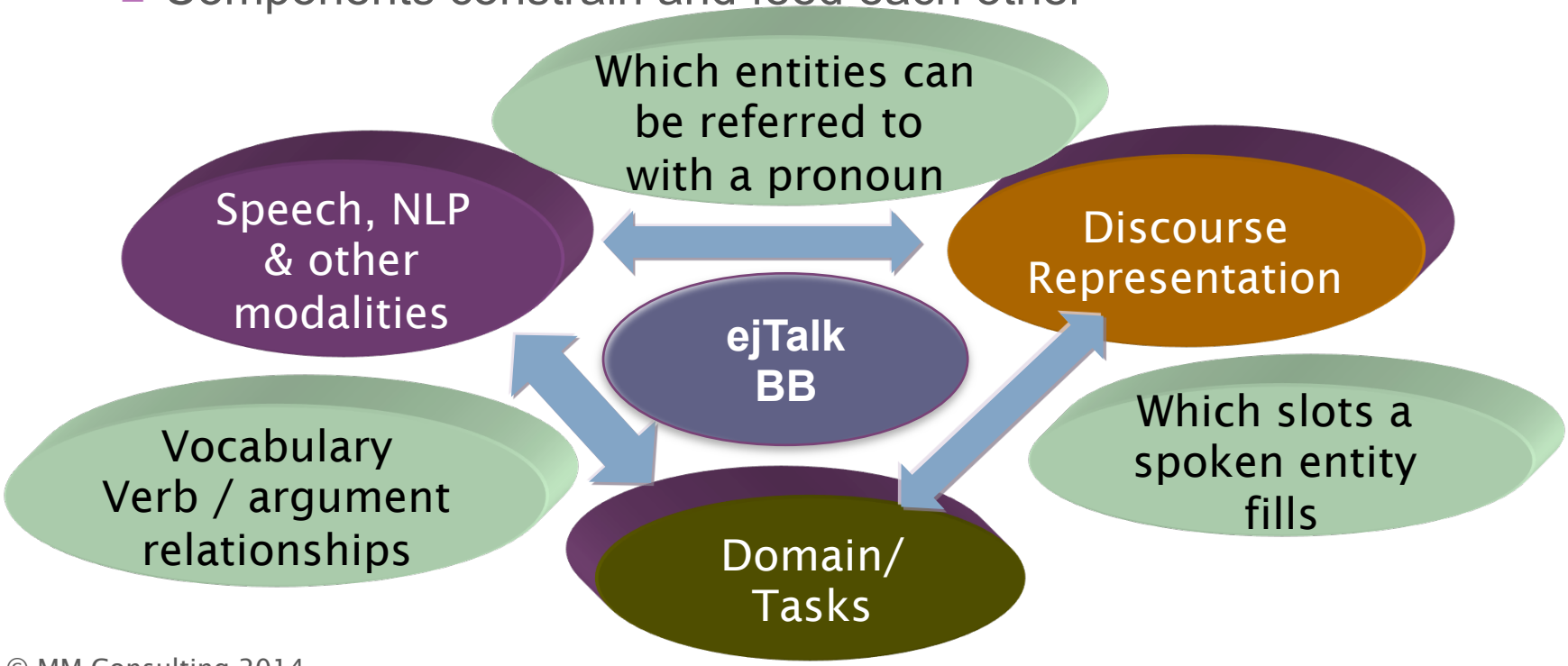
- ASR – speech to text
- NLU – identify intents and concepts
- Context – beyond the sentence or or utterance
- Virtual assistant / personal assistant / digital assistant
- Language User Interface – voice control of app
- DM – integrate ASR, NLU, Context, VA, LUI, …

+ Dialog System Architecture



+ Integrated components

- Not a strictly sequential process
  - Components constrain and feed each other



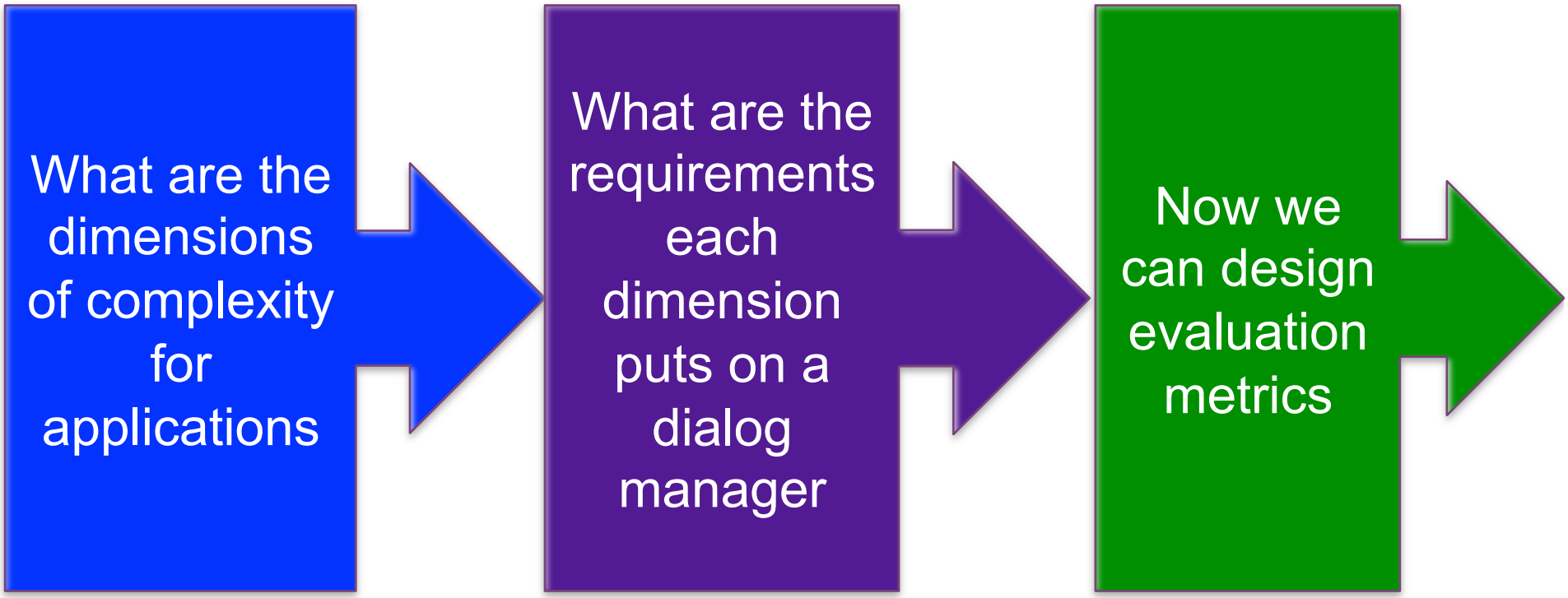
+ Assessing Dialog Managers

What metrics should we use for evaluation of dialog managers?

What are the components of dialog managers to be evaluated?

What aspects of an application need to be managed?
What does the dialog manager do?

+ Mapping the space



Dimensions of application complexity

- Task: How complex is each task, single step, requires multiple steps? How many tasks are there?
- Multitask: Are tasks completed sequentially or can there be multiple tasks “live” at the same time?
- Domain: Does the app do one thing or cover multiple domains, such as managing many different applications?
- Modality: Only speech or type/click or multiple modes for input and output?
- Connectivity and “range”: Self contained app on a single device or multiple portals into the internet of things?
- Persistence over time: Complete “transactions” in a single session or able to come back and continue on the task?

+ Dimensions of application complexity:

Number and complexity of tasks

#	Single task, Single step	Related tasks, Clear set of sequential steps	Sequential steps allowing variability	Multiple tasks with multiple steps with dependencies
	Turn on the lights	Recipe or steps to log in/set up account	Shopping, filling out an form	Book a vacation

+ Dimensions of application complexity:

Number of Domains

# Do	Single Domain	Fixed set of built in domains	Can “learn” new domains, but need to specify which domain	Open to new domains, system determines domain from discourse context
	Music player	Siri (treating web search as a single domain)	Alexa skills	My future kitchen assistant

+ Dimensions of application complexity:

Goal Driven

Only task execution	System offers “typical goal”. Simple sequence to meet goal	System can infer user’s goals and mediate competing goals	Multiple actors with shared goals and negotiated goals
Music player	Waze: “Do you want to go to work?” Goal: be at work as soon possible. Waze determines best path.	Recognizes goal from “find the nearest restaurant” and offers to make a reservation (but doesn’t offer to make a reservation at the nearest gas station)	Recognizes from big data what user might want to do next and offers up possibilities. Multiple stakeholders

+ Dimensions of application complexity:

Multimodality

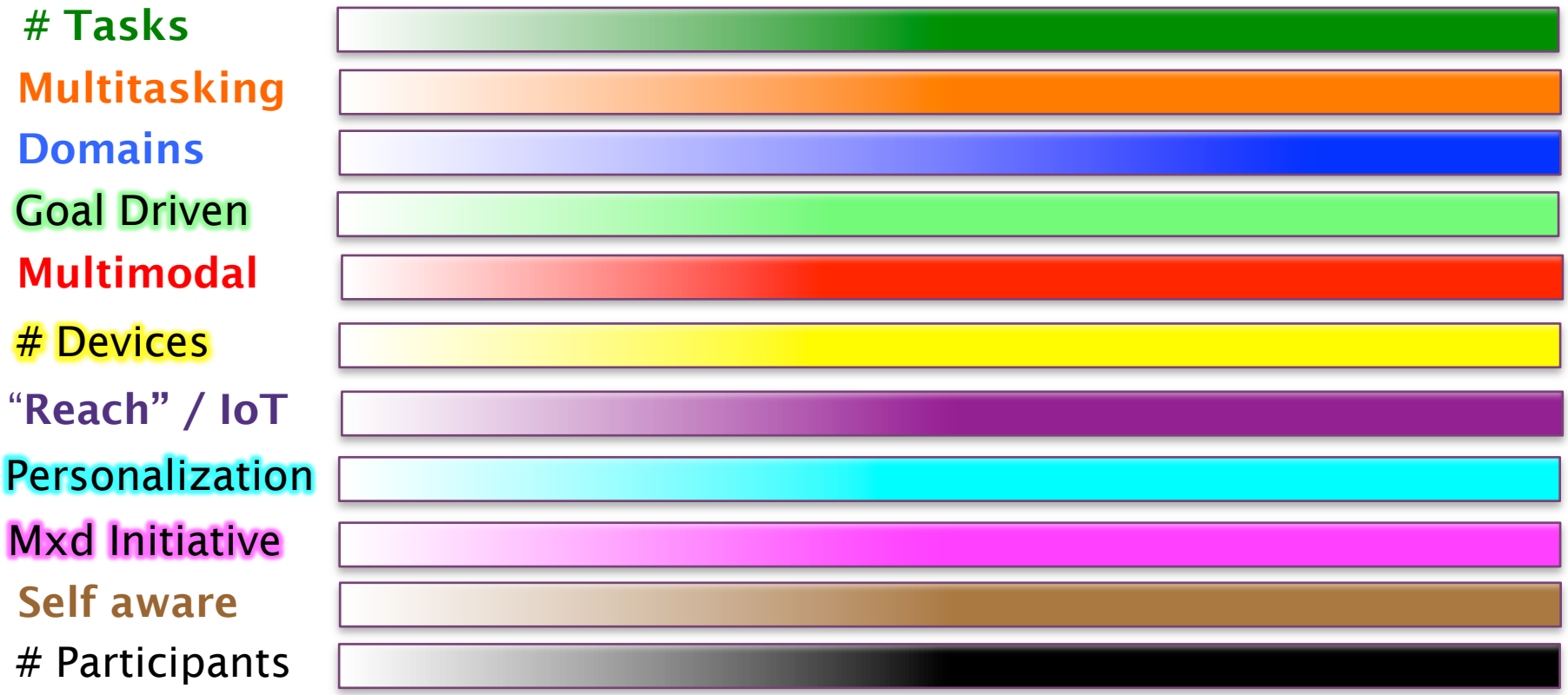
Single mode (e.g. speech or typing)	Fixed modes for specific actions	Simultaneous synergistic use of multiple modes	Automatic modality transitions depending on device, context,…
Music player: You speak, it plays music	Voice web query: You speak, it shows you the results	“I want this one” “Show me the blue top” “What’s that”? (takes a picture)	Moving from my car to my house on my phone and now info shows up on the screen.

+ Dimensions of application complexity:

Reach into the IoT

Only one "thing" on the internet	Control of devices "preprogramed"	Disambiguating many different IoT devices	Tracks goals and offers to control devices and makes recommendation
Music player: You speak, it plays music	"Turn on the living room lights" (device and name preprogrammed)	"Turn on the lights" can figure out what house/room you're in	"Printing your boarding pass. Would you like me to turn down the heat while your gone?

+ Dimensions of Applications



+ Measuring complexity

Tasks

Multitasking

Domains

Goal Driven

Multimodal

Devices

“Reach” / IoT

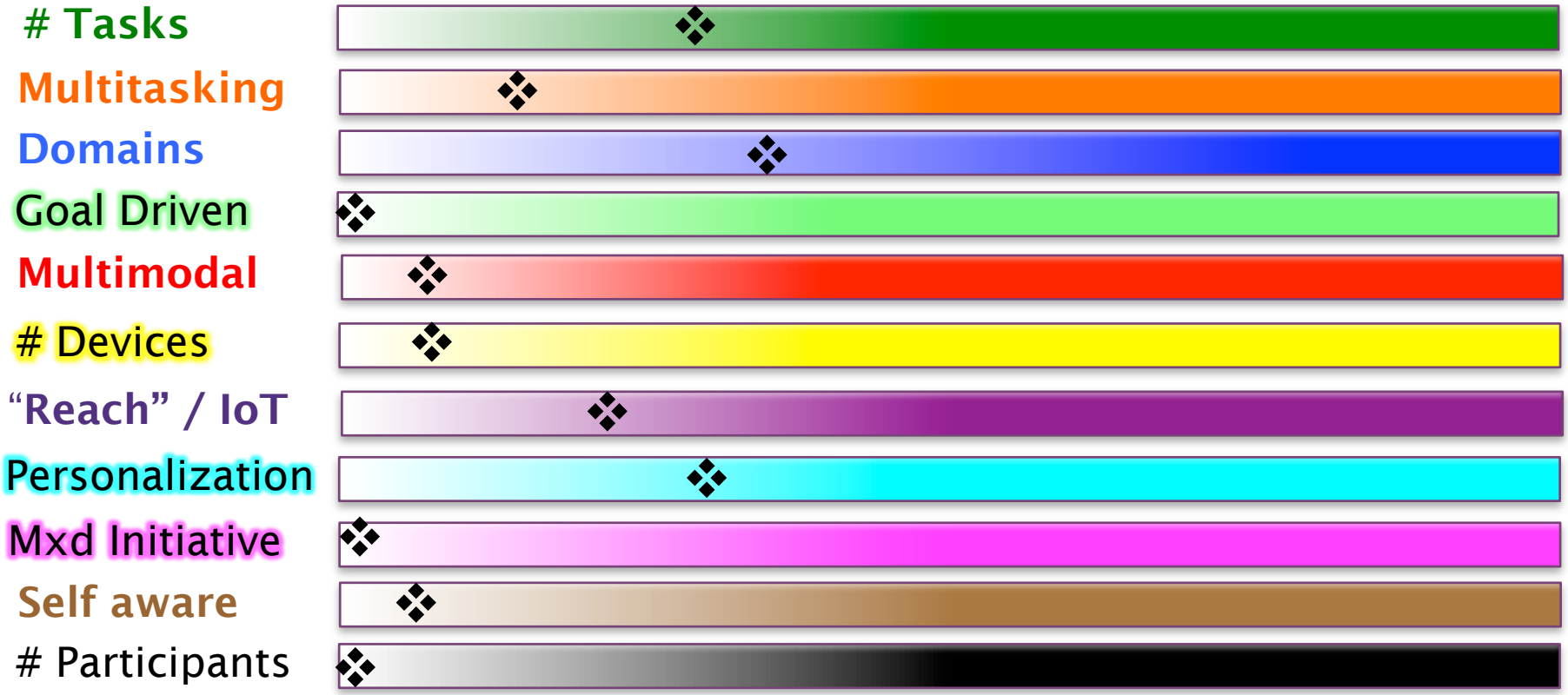
Personalization

Mxd Initiative

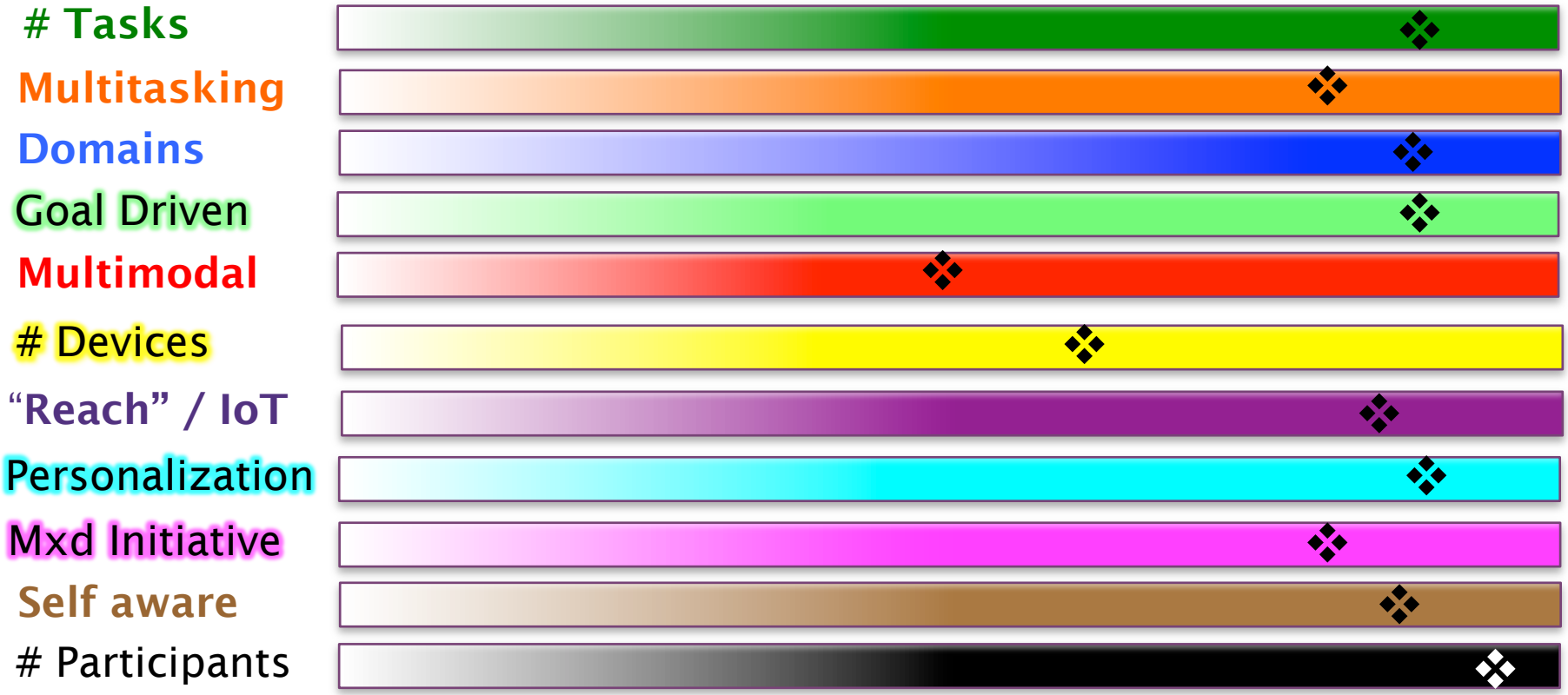
Self aware

Participants

+ The Echo on my kitchen counter



+ My Dream Echo with “Future Alexsa”



Mapping from App capabilities to DM requirements

- As Applications become more complex along these dimensions, the dialog manager needs more capabilities

Multi-step tasks	➢ Representation of the steps ➢ Tracking what is completed and what is left
Multiple domains	➢ Model of intents in each domain ➢ Set priorities of tasks in different domains
Multimodal	➢ Represent the capabilities of each modality ➢ Include visual modalities in context
Goal Driven	➢ Recognize the user's goals ➢ Select tasks that meet those goals
Reach to the IoT	➢ Model physical context ➢ Track patterns

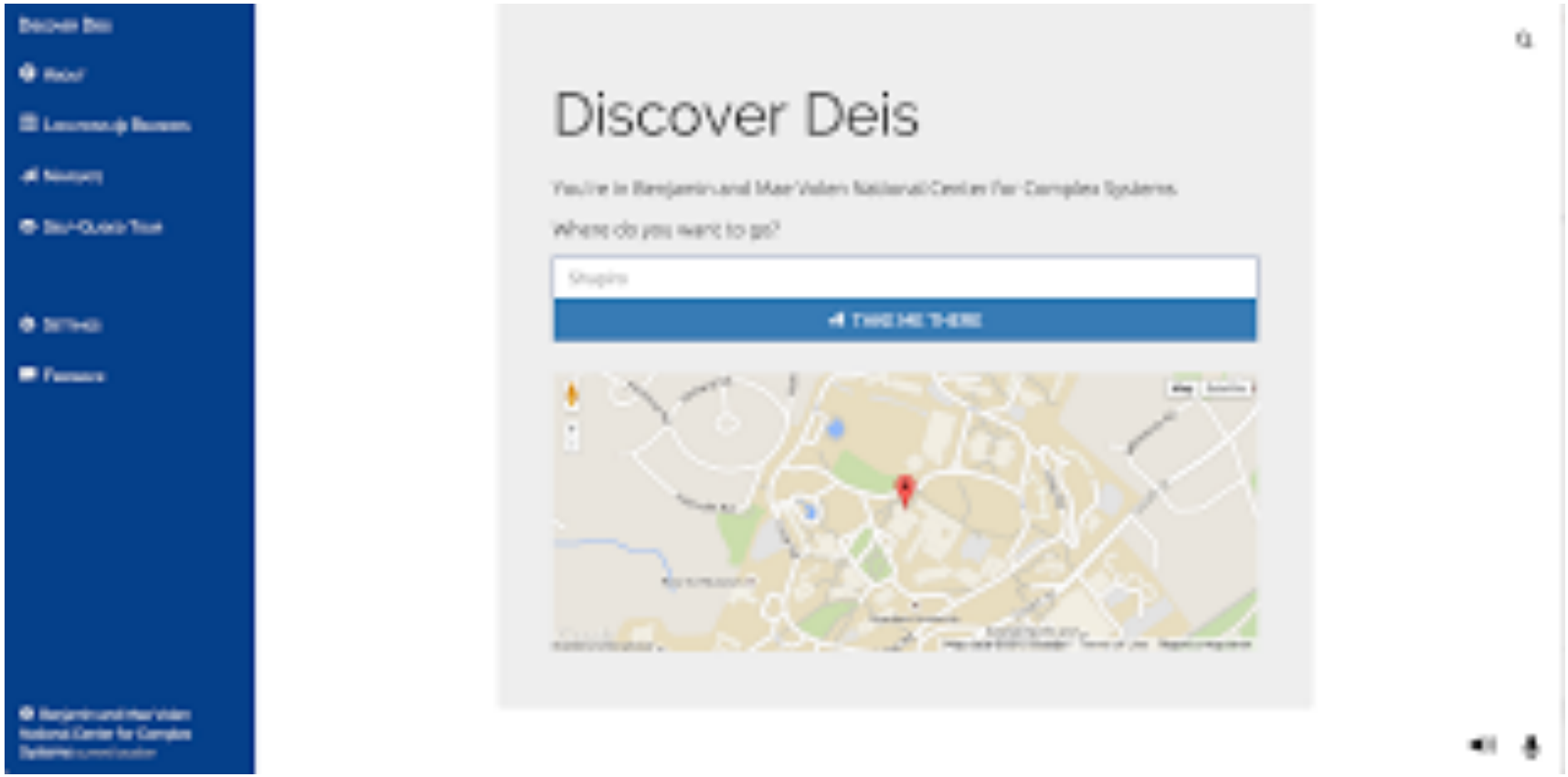
From App capabilities to Discourse requirements

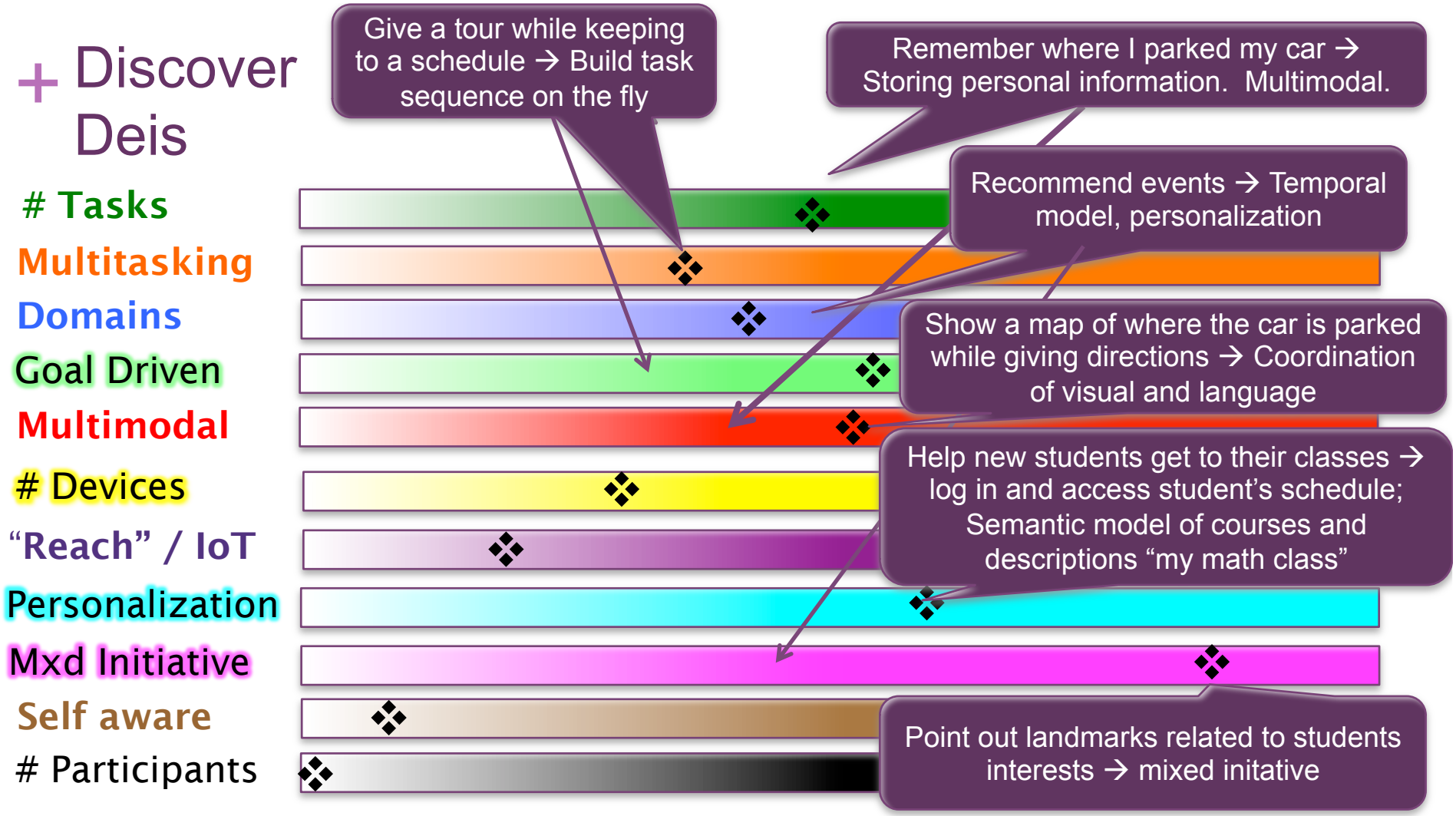
- LUI becomes more complex as well

Multi-step tasks	➢ Reference resolution ➢ Verbal clues to map steps
Multiple domains	➢ Ability to disambiguate ➢ Knowing when to clarify
Multimodal	➢ Resolving deictic references ➢ Resolving/generating verbal
Goal Driven	➢ Inferring goals from tasks ➢ Recognizing goal changes
Reach into the IoT	➢ Semantic model of what things can do ➢ Variable degrees of clarification

Discover Deis:

Web App to Navigate Brandeis




+ Discover Deis
# Tasks
Multitasking
Domains
Goal Driven
Multimodal
Devices
"Reach" / IoT
Personalization
Mxd Initiative
Self aware
Participants
Give a tour while keeping to a schedule → Build task sequence on the fly
Remember where I parked my car → Storing personal information. Multimodal.
Recommend events → Temporal model, personalization
Show a map of where the car is parked while giving directions → Coordination of visual and language
Help new students get to their classes → log in and access student's schedule; Semantic model of courses and descriptions "my math class"
Point out landmarks related to students interests → mixed initative

+ Metrics for evaluation?

- Numbers based evaluation is possible
  - Ask the research community about the dialog challenges
- Have to first focus on understanding the capabilities we want in our apps
  - Application driven measures of complexity
- Next, project capabilities to the requirements of the dialog manager
- Finally, drive implementation and evaluation from use cases
    - Remember last year “Wait, do I have enough in my budget for that?”

+ Thanks!

- **The AVIOS Advanced Dialog Group**
  - Emmet Coin
  - Marie Meteer
  - John Tadlock
  - K.W. (Bill) Scholz
  - Lorin Wilde

- The students in the **Brandeis Discourse and Dialog seminar**
  - Tuan Do, Swini Garimella, Jessica Huynh, Alex Luu, Orion Montoya, Hannah Provenza, Keigh Rim,
  - ***See me if you're hiring! Summer interns available.***