Starting and selling an AI startup

Agenda

- Background on my startup (Parakweet Labs)
- Approaches we used and challenges we faced
- Lessons learned
- General challenges startups building AI applications will face

Overview of Parakweet Labs

- Started the company in mid 2010
- Raised our first round of funding in 2011
- Team of 8 full time people when we sold
- Based in San Francisco (with 2 engineers overseas)
- Sold in early 2016

Overview of Parakweet Labs (contd).

- Raised $4m in funding in total
- Received funding from top tier Silicon Valley Angels. Some of the angels included:
  - Scott Banister (early investor in Paypal, Uber)
  - David Jeske (sold Neotonic to Google)
  - Amit Kulkarni (sold Manymoon to Salesforce)
  - Alan Braverman (co-founder of Yammer and Eventbrite)
  - Vipul Ved Prakash (sold Topsy to Apple)

Team

Kiam Choo
Co-founder, CTO

Ramesh Haridas
Co-founder, CEO

William Pearce
President, Head of Product

Ilmars Poikans
ML and NLP engineer

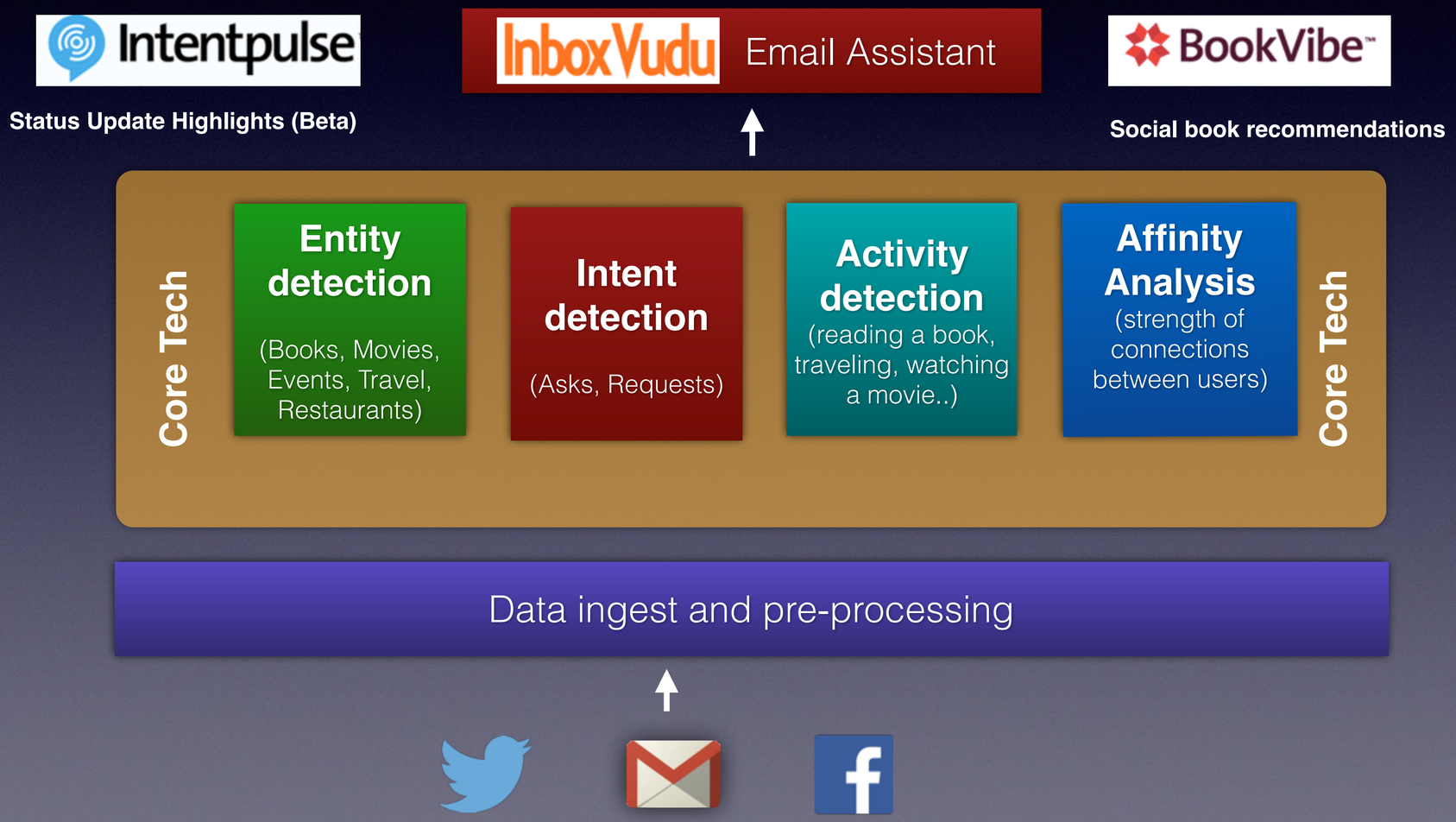
Duan Tran
ML and NLP engineer

Xiaozhi Zhang
Infrastructure Engineer

Gary Graham
Backend Engineer

Bo Ma
Data Analyst

Experimented with multiple products on the Parakweet Engine



What was InboxVudu?

Personal email assistant that used Natural Language Understanding and Machine Learning to enable:

A

- **Prioritization** - importance of emails assessed based on content and sender importance

B **Summarization** - intelligent snippets that display the essence of the message

C **Acceleration** - quick replies to shorten response cycle and boost productivity

We detected actionable emails

Requests
for you to do something
(e.g., send a file)

Proposals
for joint activities
(e.g., meeting scheduling)

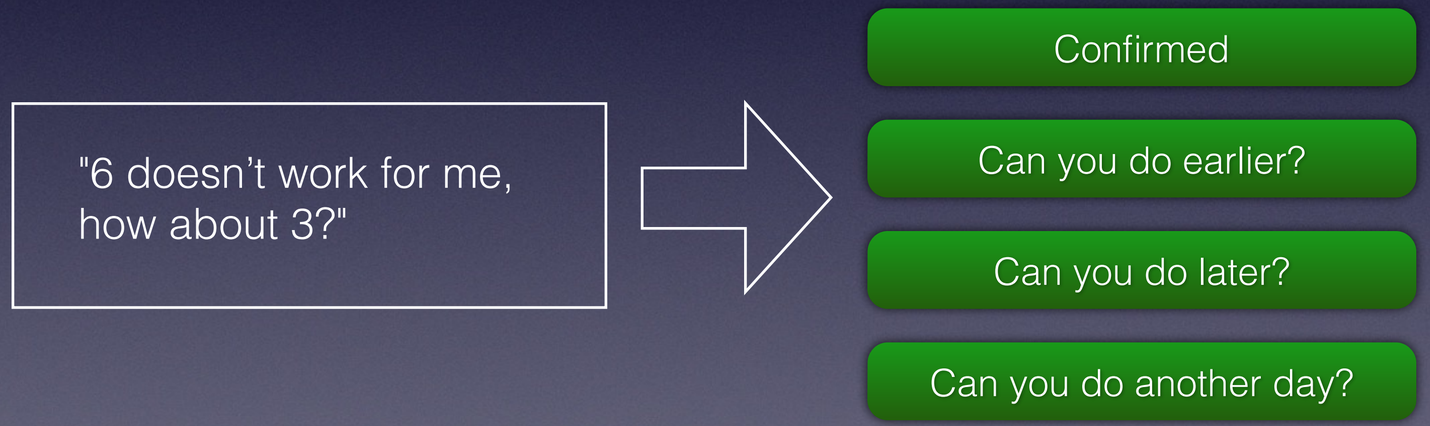
Questions
for you to answer
(e.g., What filetype do you want?)

Commits
confirming tasks or agreements
(e.g., confirming a meeting)

We would show them until they were actioned

Mobile App

- Notification of important requests on iPhone + Apple Watch
- One-tap replies. Example:



- Integration with calendar for meeting emails

Summarization

Our route to a fully-fledged intelligent assistant

Calendar Integration

Intelligent notifications based on Inbox signals (restaurant/event recommendations from friends, previous bookings)

Additional Smart Replies for Consumers

Machine mediation for meeting scheduling, booking restaurants/tickets etc

Customized Smart Replies for Enterprises

InboxVudu received excellent press coverage

Forbes

“Prioritizes your email…sign up takes 10 seconds”

GeekWire *“Uses AI to keep on top of important emails”*

TNW THE NEXT WEB *“Makes your Inbox smarter”*

TechCrunch *“Helps you prioritize important emails”*

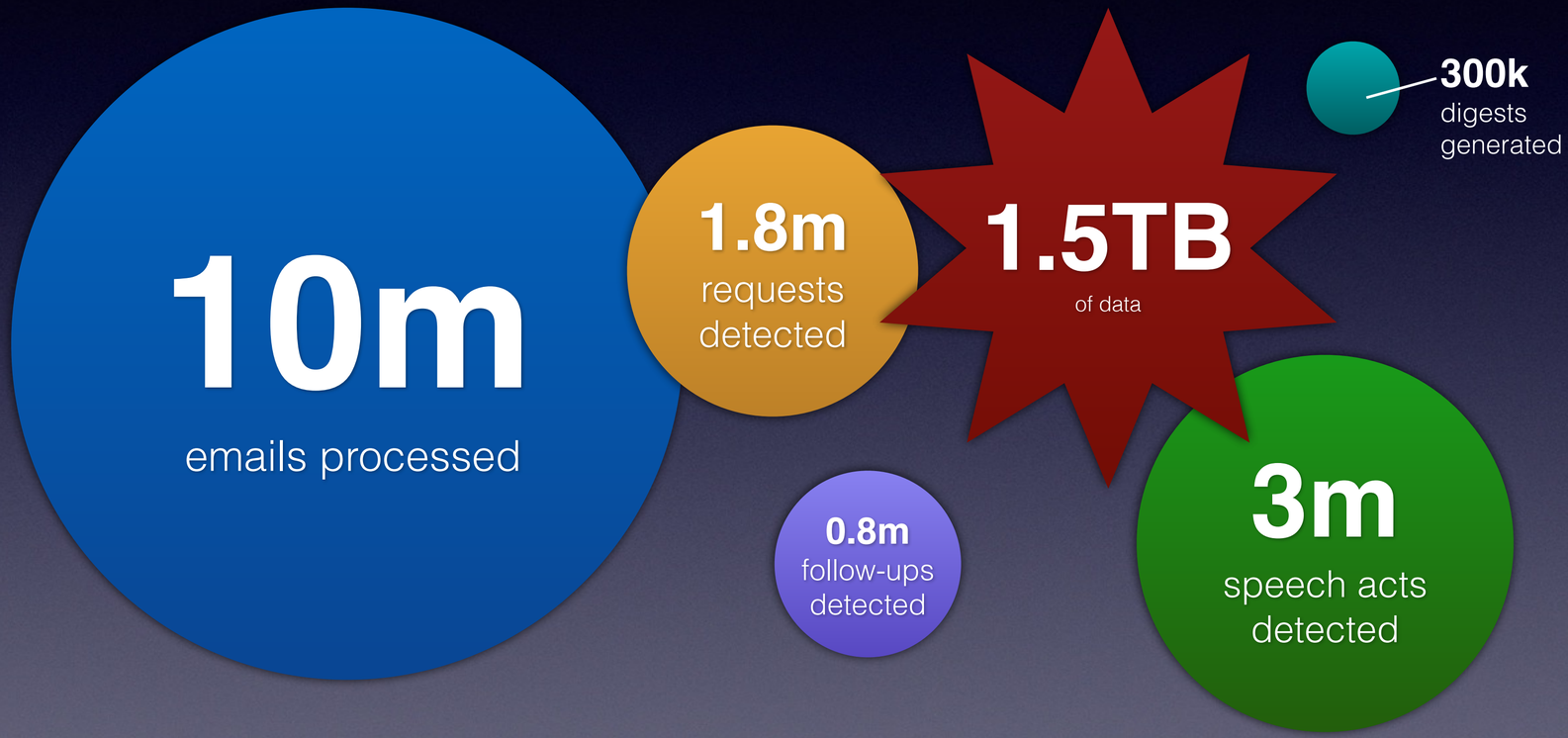
“Uses NLP to help you focus on emails that matter”

GIGAOM

Email training data

- Labeled Enron data
- Labeled 15k internal emails
- Then leveraged data from 10m emails:
  - Positive: short replied emails with no detections
  - Negative: unreplied read emails

Data advantage: we processed a large number of emails



The largest publicly available corpus of emails is only **0.5m** strong

Better than academic studies…

Email Intent Detection

- Not much work done on sentence level detection. Our F1 = 78%
  - Cohen, William W., Vitor R. Carvalho, and Tom M. Mitchell. "Learning to Classify Email into``Speech Acts''." EMNLP. 2004.

Entity Detection

- Our F1 score = 75%; top comparable academic F1 score = 54%
  - Ritter, Alan, Sam Clark, and Oren Etzioni. "Named entity recognition in tweets: an experimental study." Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2011.

Relation Extraction

- Similar to relations in Open Information Extraction, but works better for conversational text and optimized for our purposes
  - Etzioni, Oren, et al. "Open information extraction from the web."Communications of the ACM 51.12 (2008): 68-74.

Challenges - But people express themselves in many different ways

“Maybe we can have lunch or something”

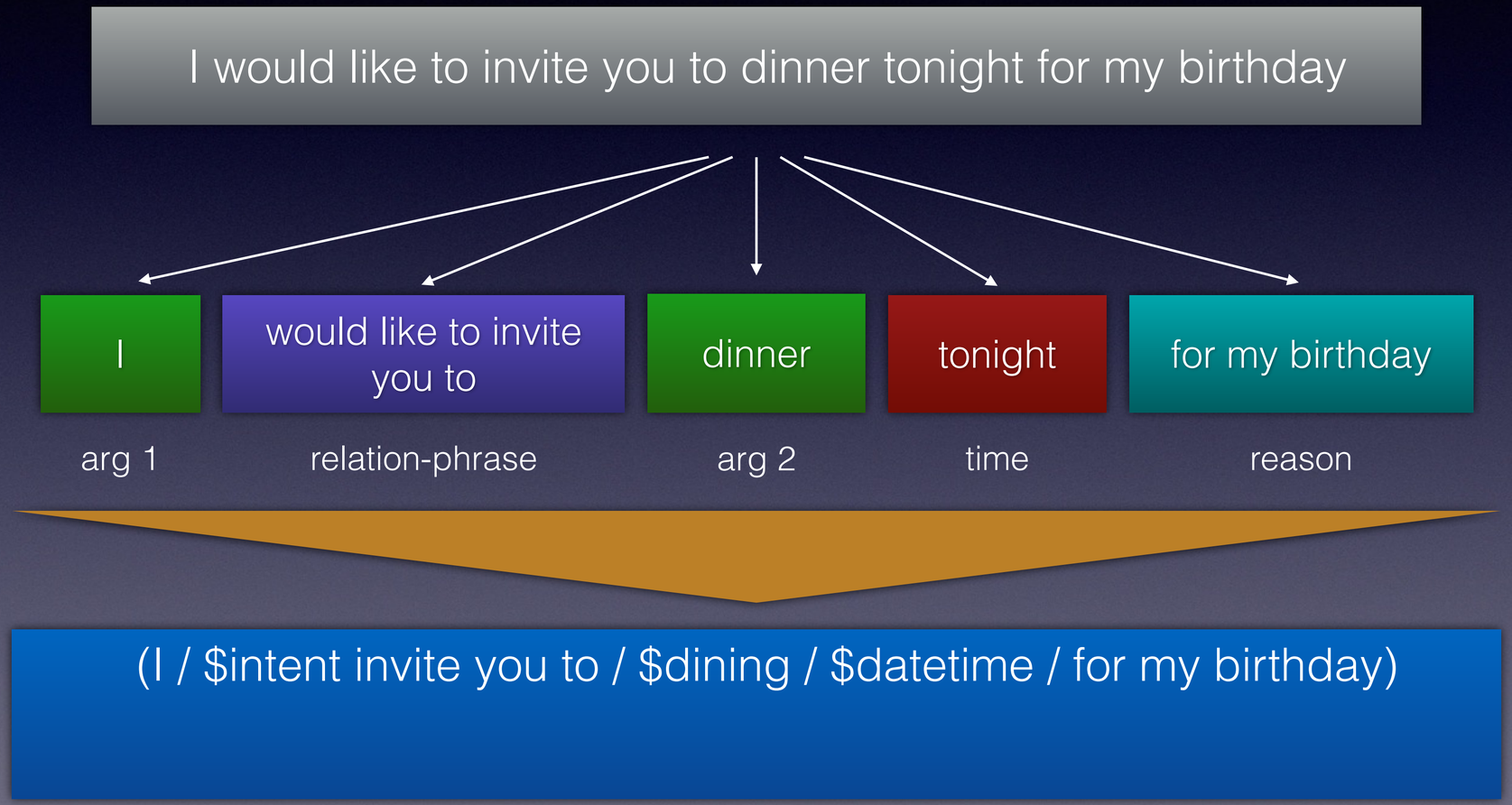
“Are you free on Friday to catch up - say 2pm?”

“Sunday, 1pm. My place. Be there.”

“I can only do Tue/Wed before work”

“Send me the address where we’re meeting”

Relation Extraction Captures Sentence-Level Structure



Approaches used

- Used a hybrid approach
  - Rules based engine
  - Deep Learning

Key lessons Learned

- The importance of clear product focus
- Measuring improvement

Key challenges faced

- Getting training data is hard for emails
- Getting clean labeled data was extremely tedious

General challenges startups will face

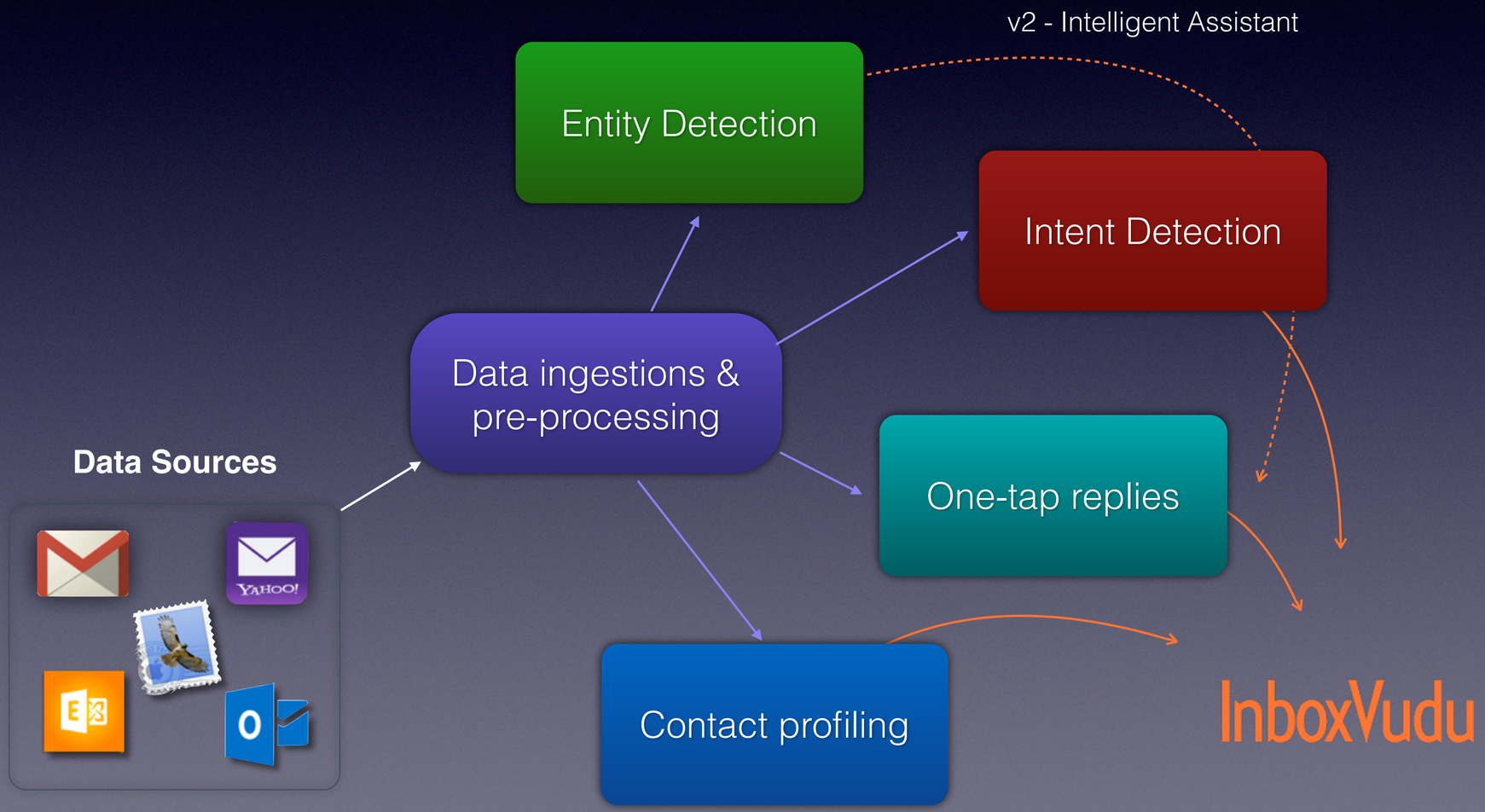
- Getting data
- Labeling data
- Presenting your system

Exiting

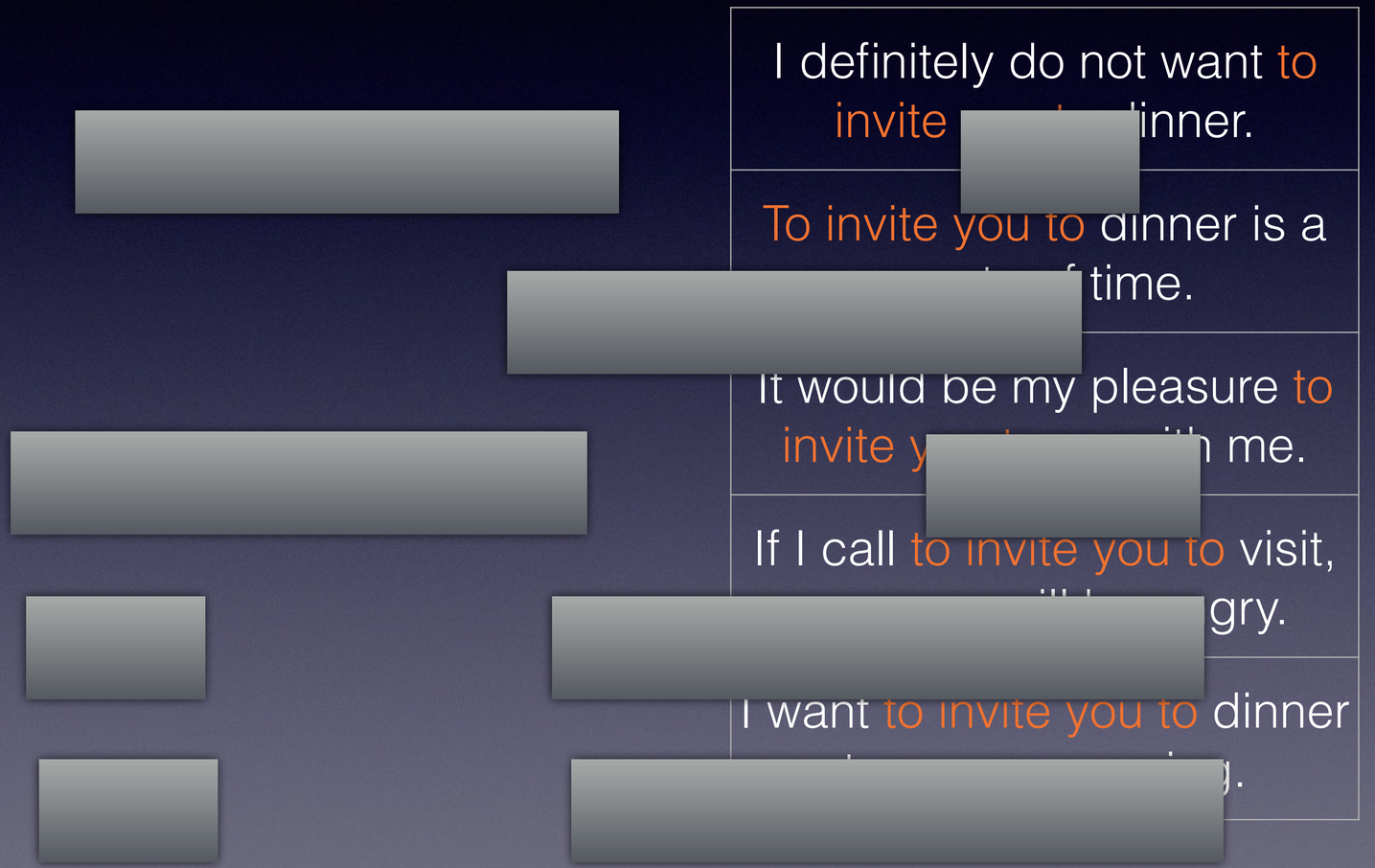
- Good exits are hard!
- Even harder if your product does not have strong traction!

Appendix

Parakweet NLP engine key components



N-grams are Poor Features for Short Text like Sentences



Relation Patterns as Features

/PRP/(me|us) *shall ask ? to give $connect* you

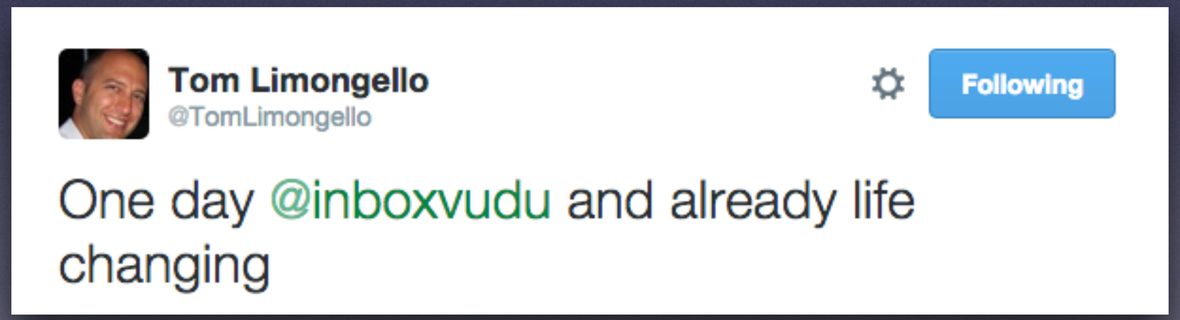
/PRP/(me|us) *shall be glad to join /PRP/you* $dining

$you *can $connect* */PRP/(me|us)*

Please do (n’t|not) hesitate to contact me

Thousands of patterns mined

Great user feedback



Very strong engagement across all products