CrowdFlower

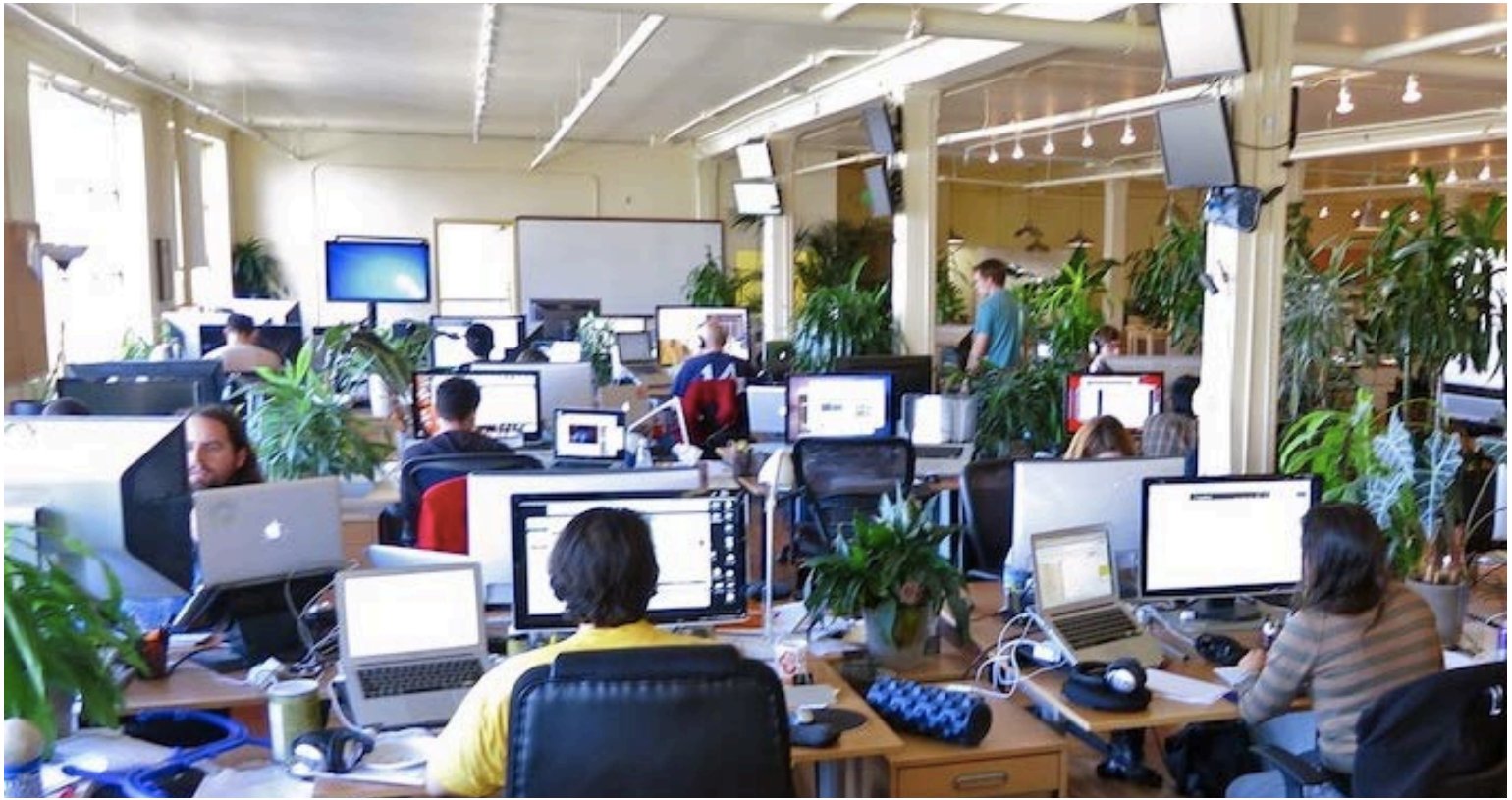
機械学習を実際のビジネスに適用する

Lukas Biewald

FEBRUARY 8, 2017

Proprietary and Confidential - Do Not Distribute

CrowdFlower



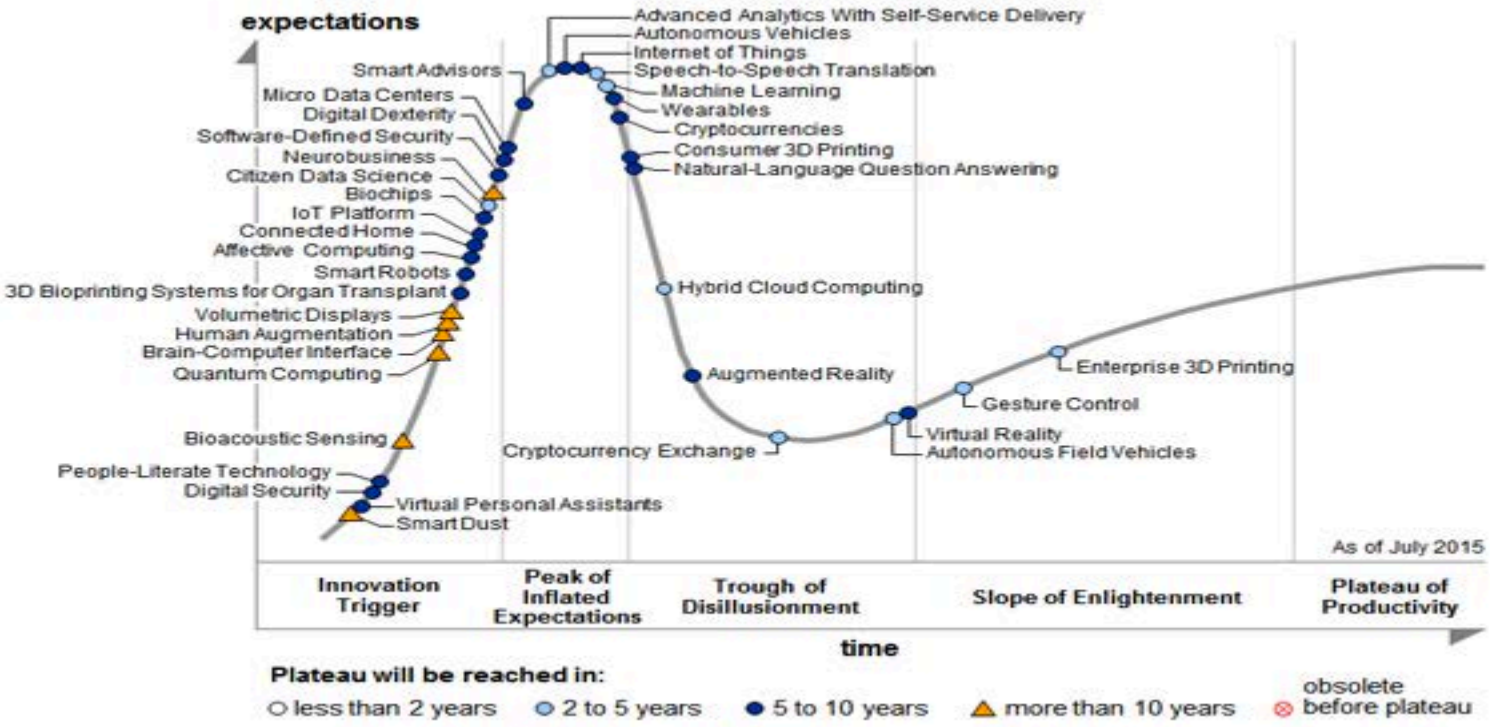
Proprietary and Confidential - Do Not Distribute

最初の仕事 (2004)

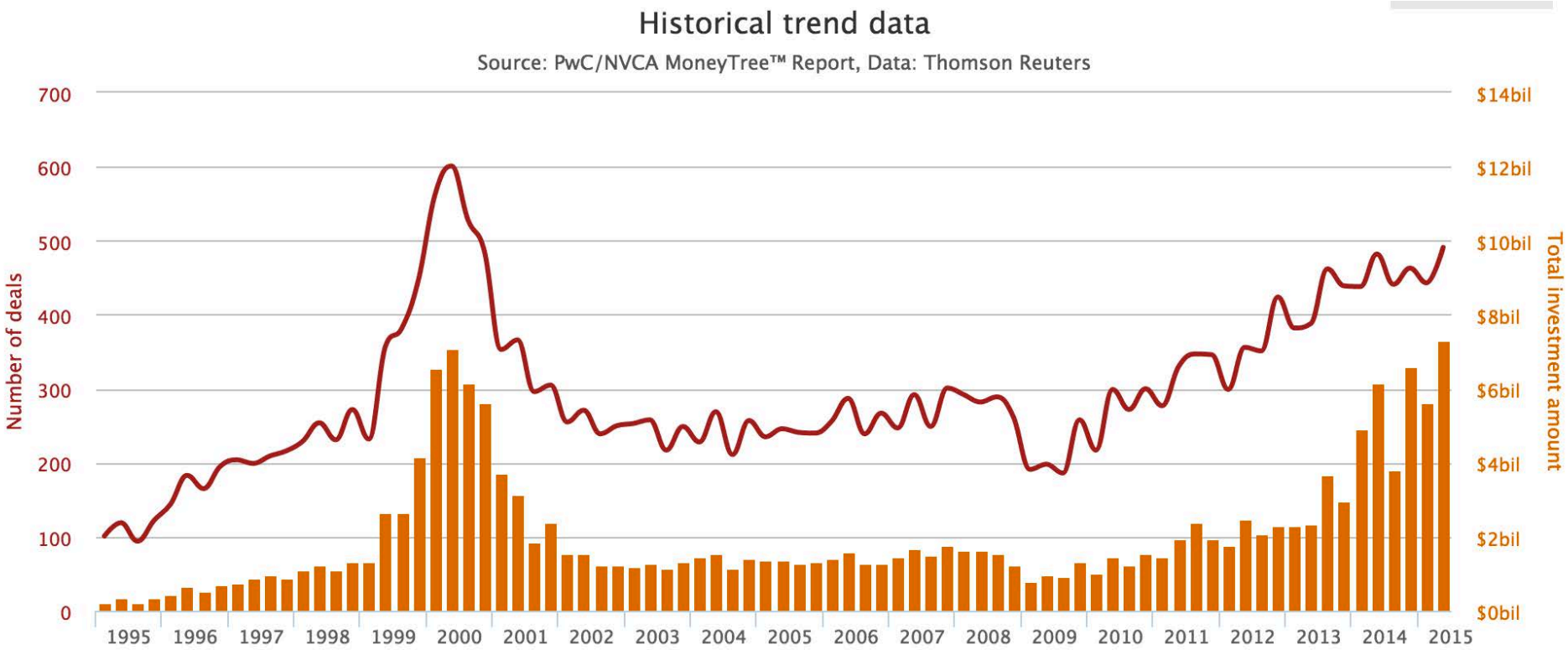
ホームページに設定する
ヘルプ
Yahoo! BB
きっず版
アプリ版
トラベル
ヤフオク!
ショッピング
YAHOO! JAPAN
プレミアム
カード
メール
カテゴリー一覧
サイトの登録
無料ID活用
ウェブ 画像 動画 辞書 知恵袋 地図 リアルタイム 一覧
検索
インフルエンザ予防 感染経路などをチェック
スノーボード選びのポイントをタイプ別に紹介
ワイモバイル契約で最大1万5千ポイント進呈
主なサービス 一覧
ショッピング
ヤフオク!
LOHACO
旅行、ホテル予約
一休.com
ニュース
天気
スポーツナビ
ファイナンス
テレビ
GYAO!
ゲーム
Y!モバゲー
地図
路線
食べログ
求人、アルバイト
ニュース 経済 エンタメ スポーツ
0時29分更新
NYダウ 史上初の2万ドル突破
豊洲 年100億円の赤字と試算
横浜市長 教育長の発言謝罪
「自白強要」崔被告が絶叫
アジア大会 アパの書籍撤去へ
北京リレー 日本の銅が銀に
アニソン歌手Ray 活動休止へ
世界仰天ニュース 火曜9時に
もっと見る 記事一覧
そっくりさん現る
1月25日20時1分配信
アフロ
カラダにピース CALPIS
8週間で血管年齢が若返る!?
血管年齢の若返りや、血管のしなやかさの改善が期待できる特別な成分とは？
詳しくはこちら
手軽に食べられる、人気のグルメ商品
男が選ぶカップラーメン
お肉系の缶詰ランキング
即席カレー、売れ筋は？
あると便利、冷凍パン
ログイン
IDでもっと便利に[新規取得]
ログイン履歴
登録情報
メール －メールアドレスを取得
2017年1月26日(木)
今日の天気 (小田原)
0%
11℃/ -4℃
明日の天気 (小田原)
40%
14℃/ 3℃
雨雲レーダー
伊勢原市

ガートナーのハイプ・サイクル

機械学習



ベンチャーキャピタル投資



Proprietary and Confidential - Do Not Distribute

データ科学投資

Hilary Mason Leaves Bitly To Join VC Firm Accel As Its First Data Scientist In Residence

IBM Launches A $100 Million VC Fund For Watson-Based Tech Startups

Zetta Venture Partners Closes $60M Fund to Back Data-Analytics Startups

With A Focus On Big Data Startups, Data Collective Redefines Thematic Investing And Debuts Its Third Fund

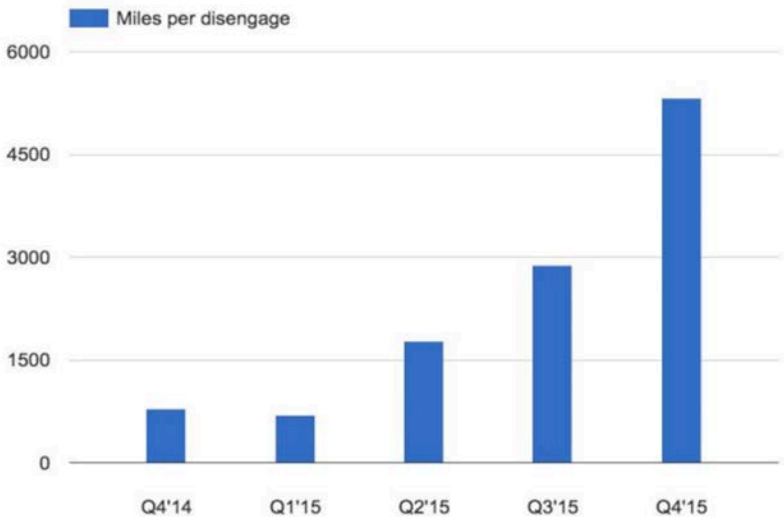
Workday Launches Data Science Venture Fund

Workday Ventures plans to invest in 10 to 12 seed and early-stage companies, and gain access to top data science talent

大きな勝利

Proprietary and Confidential - Do Not Distribute

自律走行車


Miles per disengage
6000
4500
3000
1500
0
Q4'14
Q1'15
Q2'15
Q3'15
Q4'15

機械学習プラットフォーム

Proprietary and Confidential - Do Not Distribute

機械学習を活用するCrowdFlower顧客

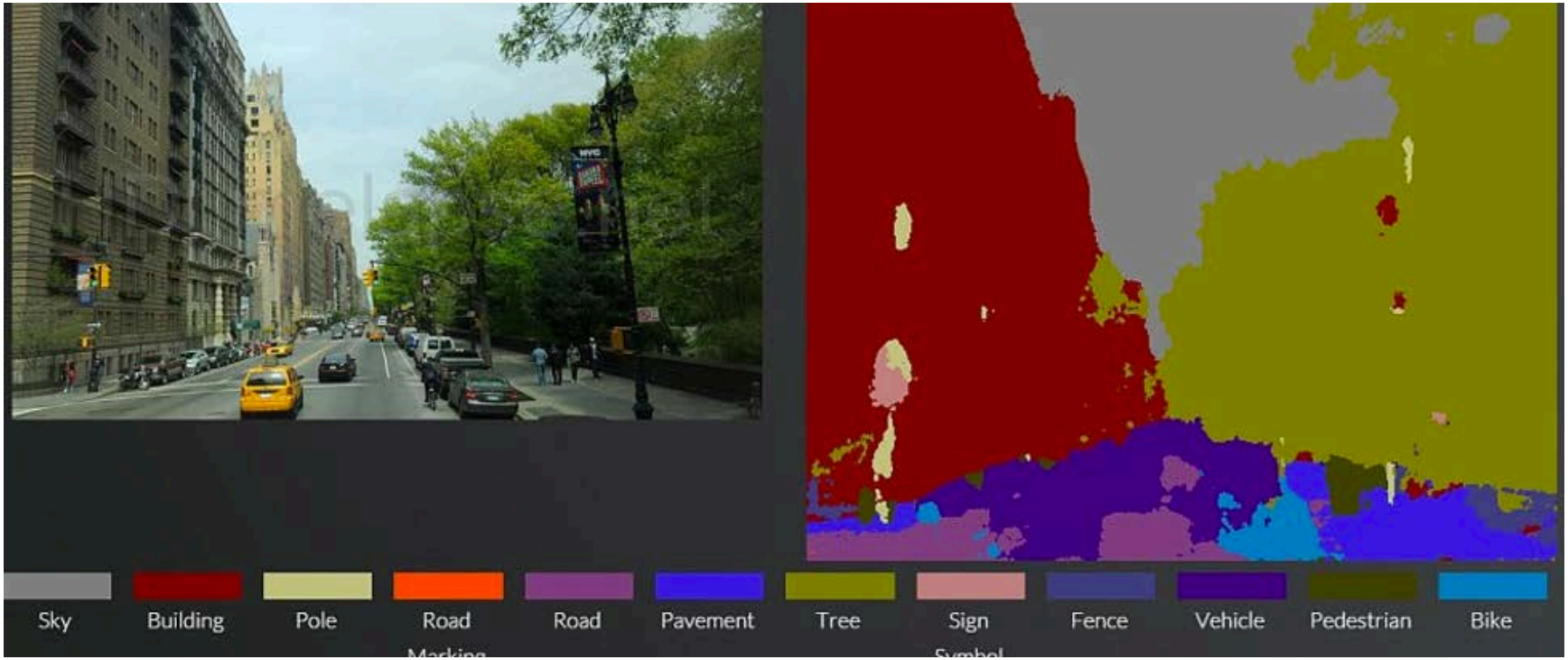
オンラインショッピング検索の関連性

顧客サポートチケット

Proprietary and Confidential - Do Not Distribute

ソーシャル　メディア　リスニング

コンピュータビジョン



Proprietary and Confidential - Do Not Distribute

機械学習はアルゴリズムだけではありません

AI = TD + ML + HITL

Training Data
トレーニングデータ

アルゴリズム

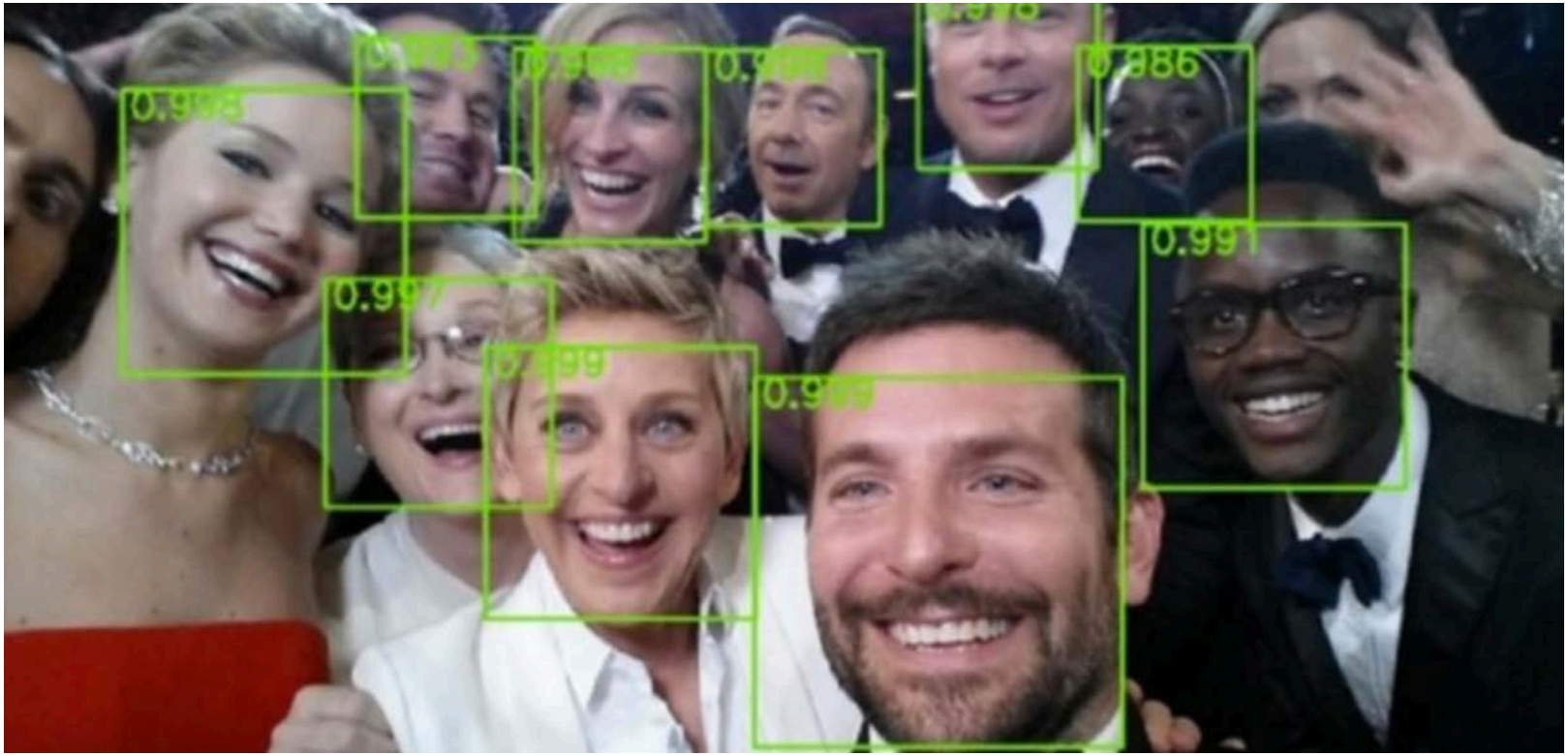
Human-in-the-loop
人間参加型

Proprietary and Confidential - Do Not Distribute

アルゴリズムは人よりスマートではありません…

アルゴリズムは人よりも忍耐強い

Facebookのトレーニングデータ



グーグル翻訳のトレーニングデータ

Abstract Type **Organizational-Scheme Type**

<abs id=A006 type=T001 str=S003>
<issue id=J001>Proc. 28th Annual Meeting of the ACL, 1990.</issue>
<title id=EA006-t>MEMORY CAPACITY AND SENTENCE PROCESSING</title>
<title id=JA006-t>記憶容量と文処理</title>
<author>Edward Gibson</author>
<keyword>memory capacity, sentence processing, working memory, garden-path effects, short term memory</keyword>

Sentence Role

<s id=EA006-1.1 role=R010>The limited capacity of working memory is intrinsic to human sentence processing, and therefore must be [addressed@@NP1 address NP2@@] by any theory of human sentence processing.</s>
<s id=JA006-1.1 role=R010>作業記憶域の容量が限られていることは、人間の文処理に内在していることであり、従って、人間の文処理に関するすべての理論によって取り組まれなければならない。</s>

Verb Complementation Pattern

<s id=EA006-1.2 role=R020>This paper [gives@@NP1 give NP2@@] a theory of garden-path effects and processing overload that is [based@@NP1 base NP2 on_PP@@] on simple assumptions about human short term memory capacity.</s>
<s id=JA006-1.2 role=R020>本論文は、人間の短期記憶容量に関する単純な前提に基づく、ガーデン・パス効果と処理過負荷についての理論を述べる。</s>

</abs>

Proprietary and Confidential - Do Not Distribute

オンラインショッピング検索のトレーニングデータ

Query
Understand Intent:

The search term is:

laptop lenovo

Google Search for laptop lenovo

Result
Product Title:

Lenovo ThinkPadT420 Intel Corei5 2.5GHz 4GB 750GB 14in Wi-Fi DVDRW CAM Windows 7 Professional (64-bit) (Refurbished)

$354.99

Result
Product Image:

Product Page

How well does this result match the query?

Off Topic
- The intent of the query was not matched
- The results are irrelevant to the search query

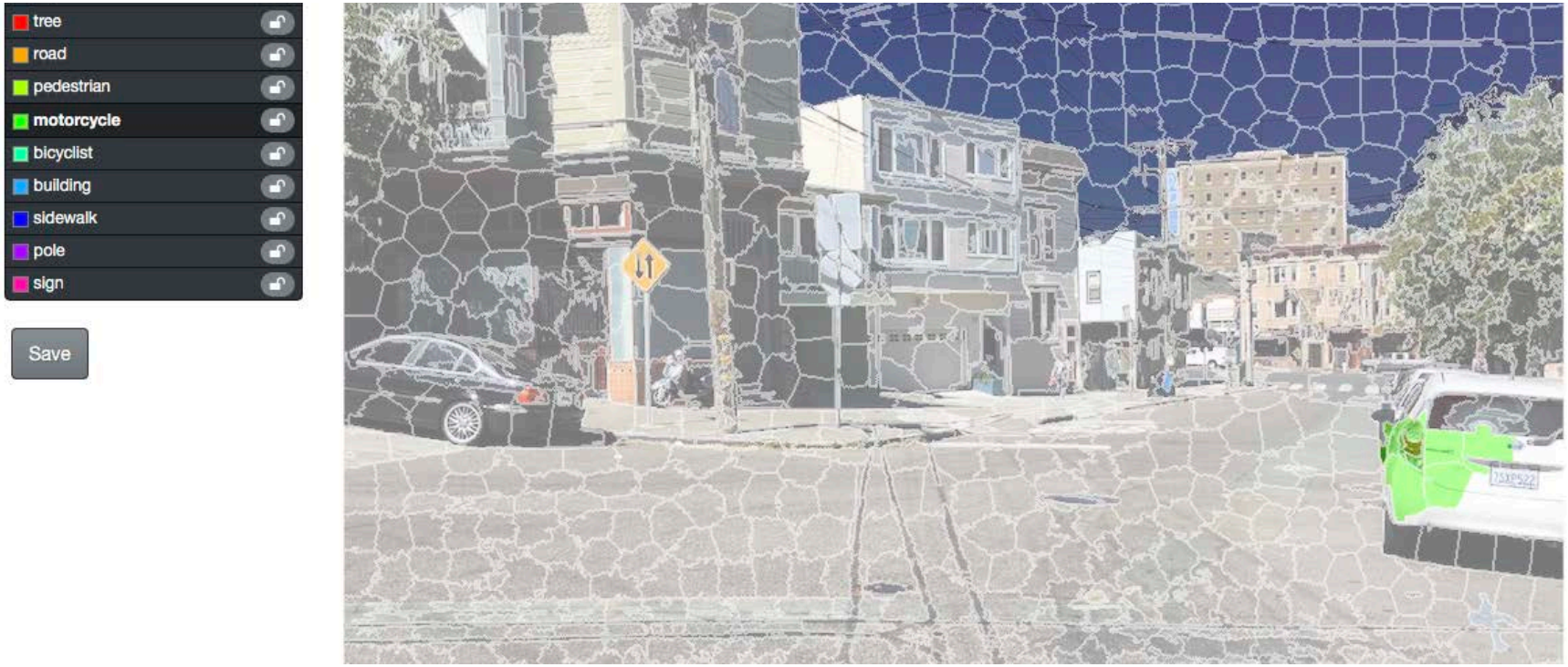
Acceptable
- The intent of the query is poorly matched
- The result is somewhat related to the query, but it not a good match

Good
- Matches most of the query intent - or the most important part of the query.
- Technically, all parts of the intent are satisfied but result doesn't provide a full, clear and complete answer to the search.

Excellent
- The query intent is clearly satisfied. This is exactly the product I was looking for
- Result is high quality
- Specifics of the Query appear in the Result

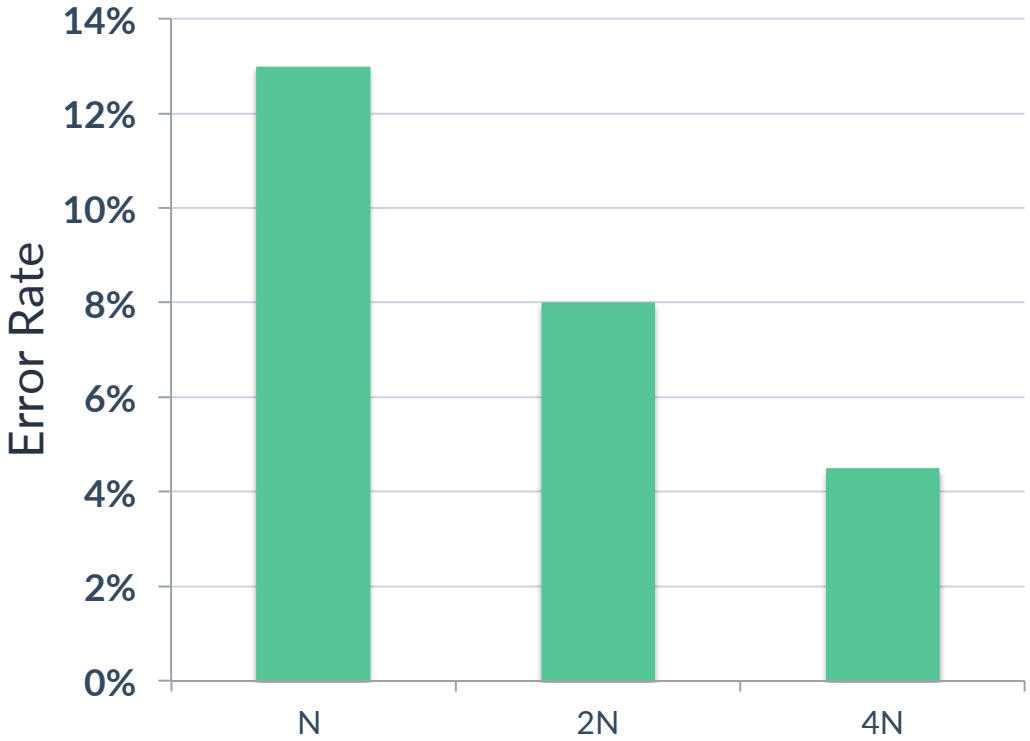
Proprietary and Confidential - Do Not Distribute

自律走行車のトレーニングデータ



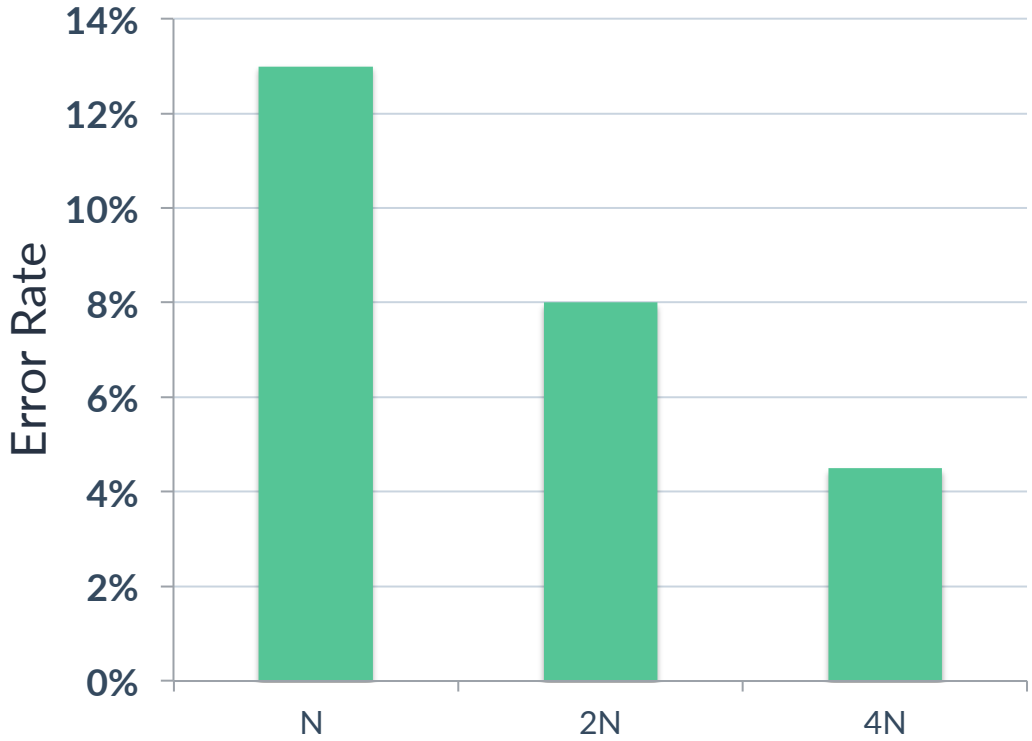
Proprietary and Confidential - Do Not Distribute

機械学習にはトレーニングデータが必要



Proprietary and Confidential - Do Not Distribute

アルゴリズムはまだ間違う…?



Proprietary and Confidential - Do Not Distribute

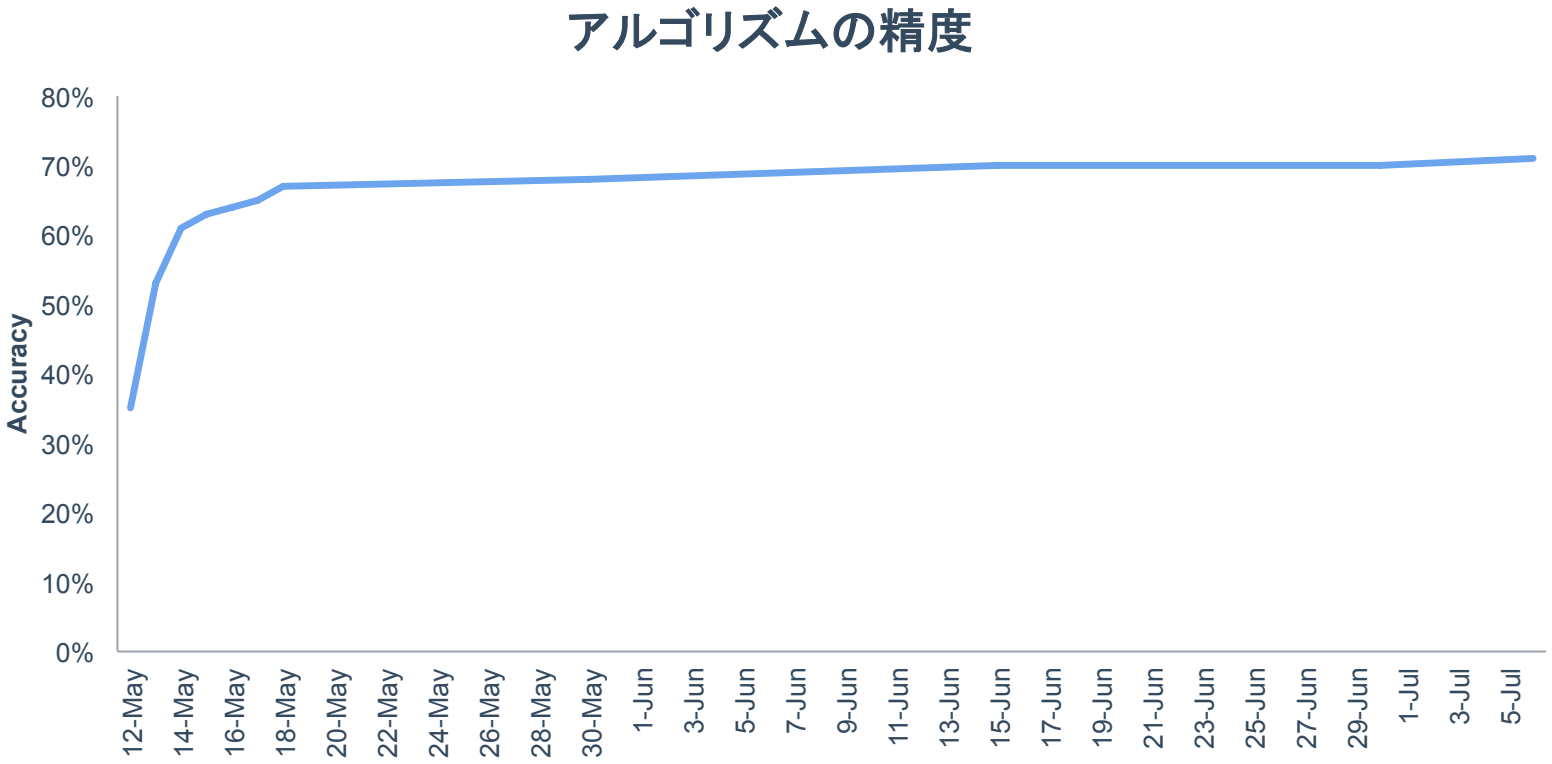
時には99%の精度は素晴らしいです…

時には99%の精度は致命的です…

Proprietary and Confidential - Do Not Distribute

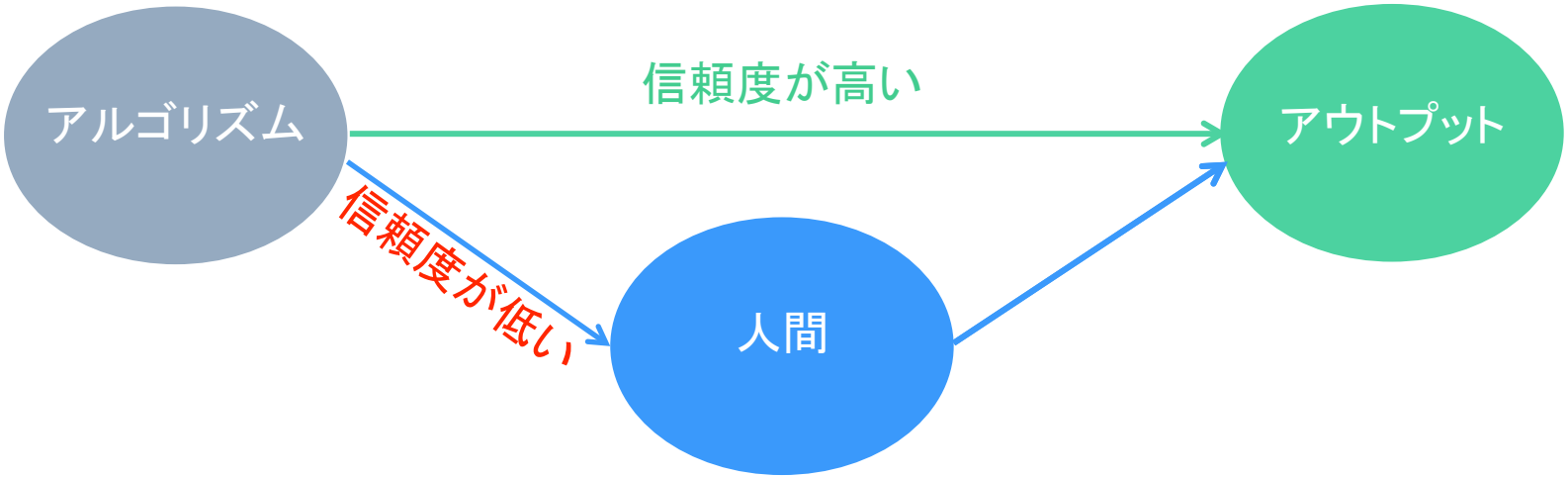
時にはアルゴリズムを改善するのは簡単

時にはアルゴリズムを改善するのは難しい

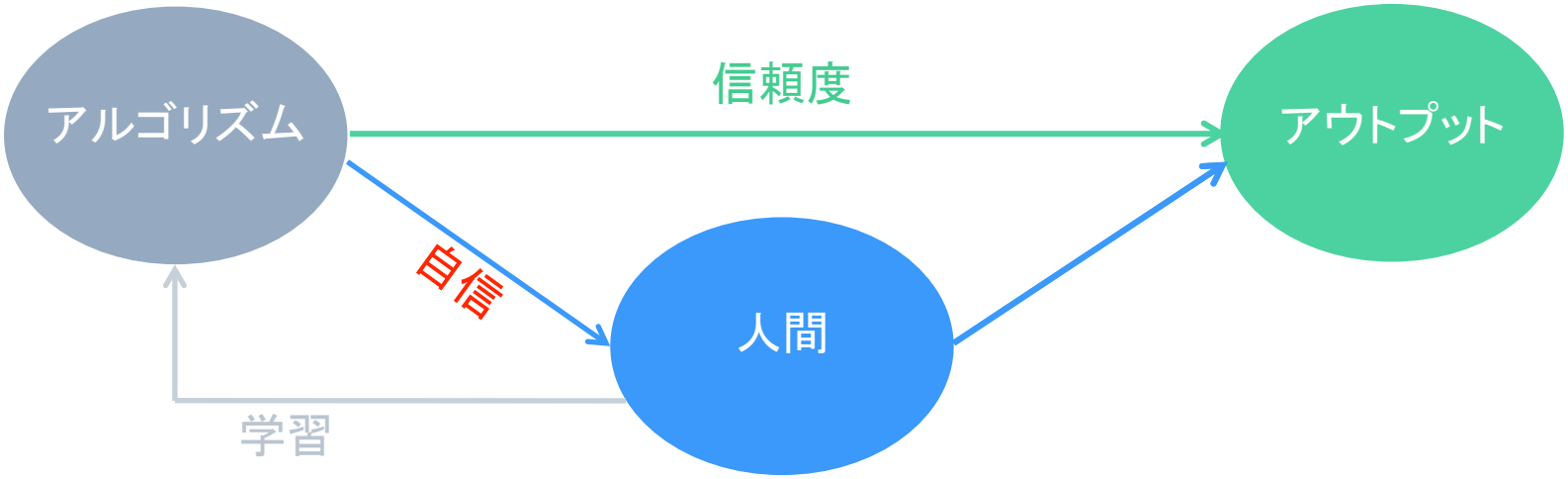


人間参加型

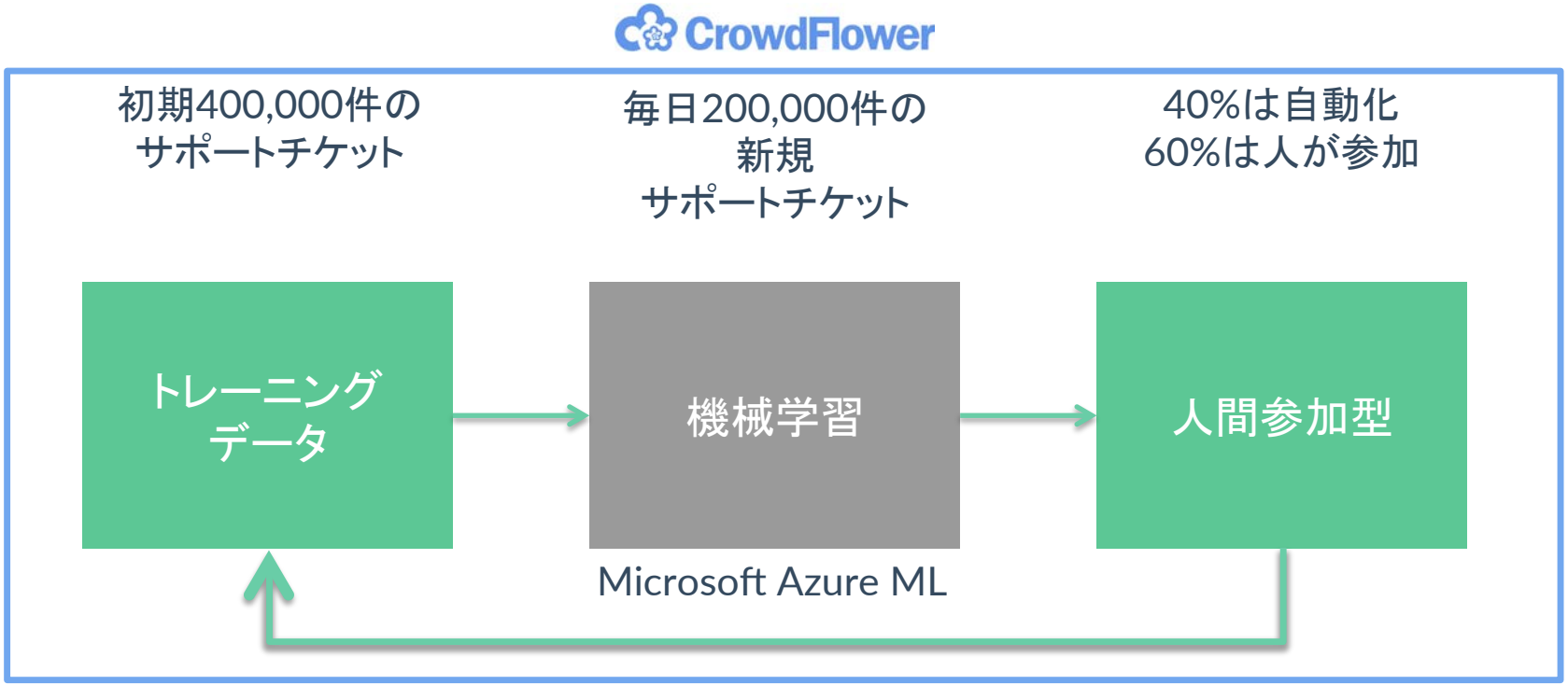
人間参加型



人間参加型



通常のCrowdFlowerの例



Proprietary and Confidential - Do Not Distribute

米国の郵便局 (1982)

CrowdFlower

ありがとうございました!!

Lukas Biewald
@L2K

Proprietary and Confidential - Do Not Distribute