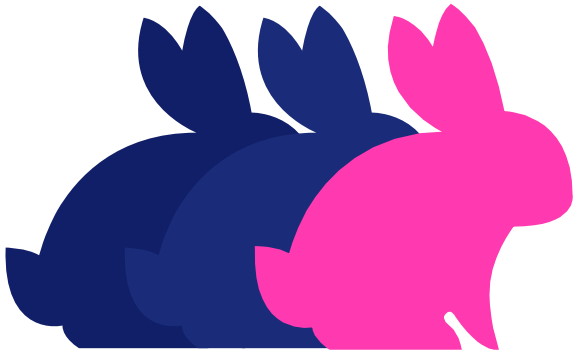
# Real World Reinforcement Learning with Vowpal Wabbit



# Agenda

- What is Vowpal Wabbit?
- New developments and what's in the pipeline
- What are Contextual Bandits?
- Real world usage of Contextual Bandits and Vowpal Wabbit
- VW + Python
- VW + Python + Contextual Bandits

# Who are we?

John Langford

Rodrigo Kumpera

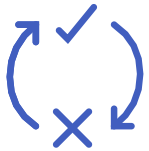
Marco Rossi

Markus Cozowicz

Y

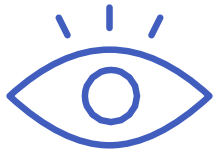
Yann Stadnicki

Jack Gerrits



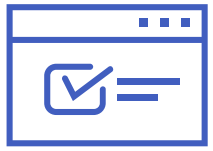
## Reinforcement learning

Contextual bandits
Learning to search
Sample efficient
Solve real world problems



## Supervised learning

Extreme classification



## Interactive learning

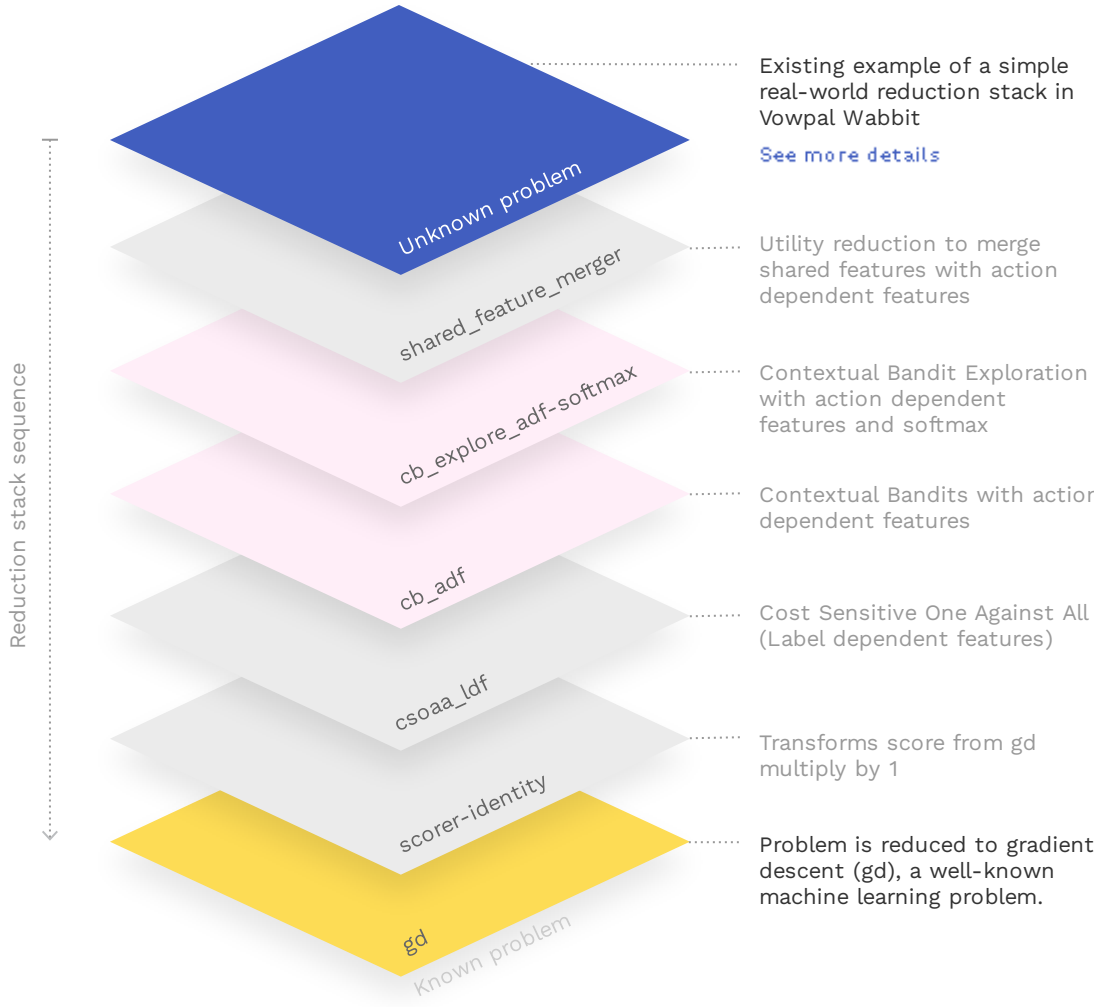
Online
Active learning

## Efficient learning

Speed
Scalability
Sparse by default

## Versatile learning

Input format
Feature hashing


Unknown problem
shared_feature_merger
cb_explore_adf-softmax
cb_adf
csoaa_ldf
scorer-identity
gd
Known problem
Reduction stack sequence
Existing example of a simple real-world reduction stack in Vowpal Wabbit
See more details
Utility reduction to merge shared features with action dependent features
Contextual Bandit Exploration with action dependent features and softmax
Contextual Bandits with actior dependent features
Cost Sensitive One Against All (Label dependent features)
Transforms score from gd multiply by 1
Problem is reduced to gradient descent (gd), a well-known machine learning problem.

# Use at Microsoft

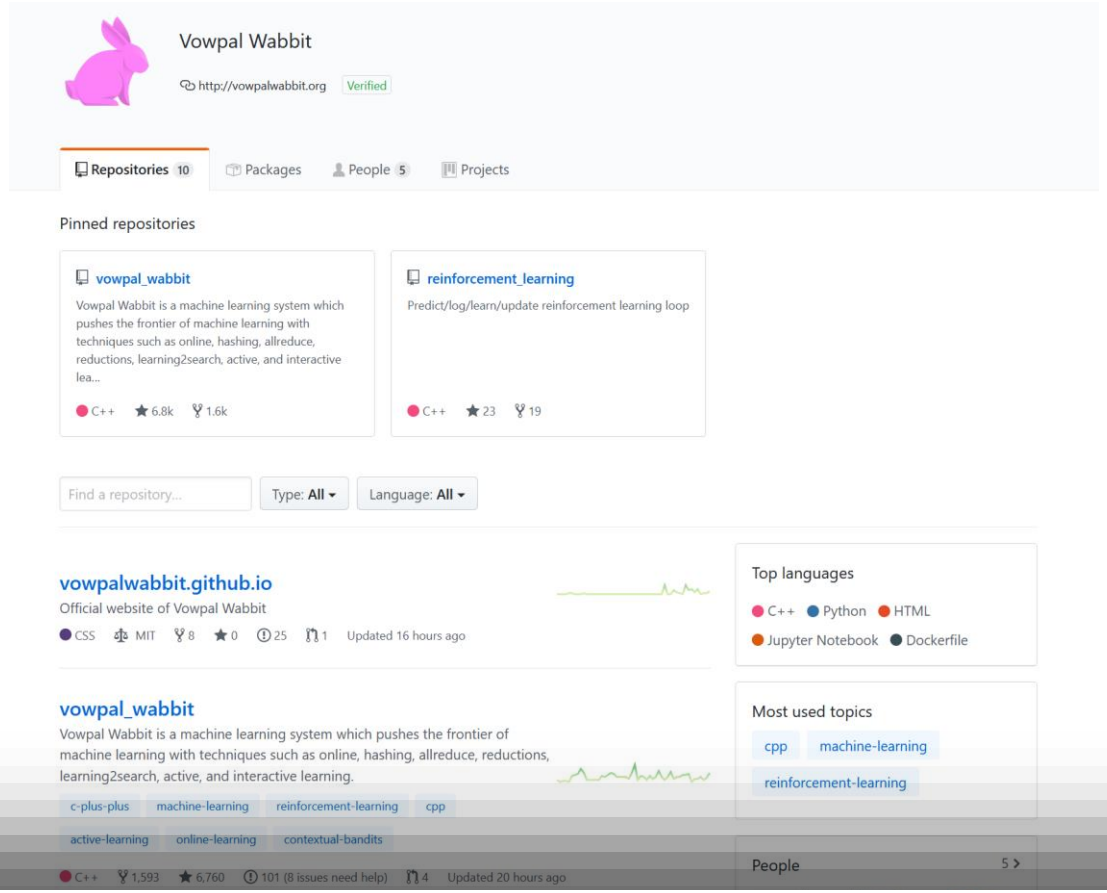
- Powers the recently released Personalizer Cognitive Service
  - Contextual bandits for personalization scenarios
- Internally for personalization and optimization scenarios
- Vehicle for experimentation and implementation of bleeding edge research
  - Meaning there is regularly new research implemented

# What’s new?

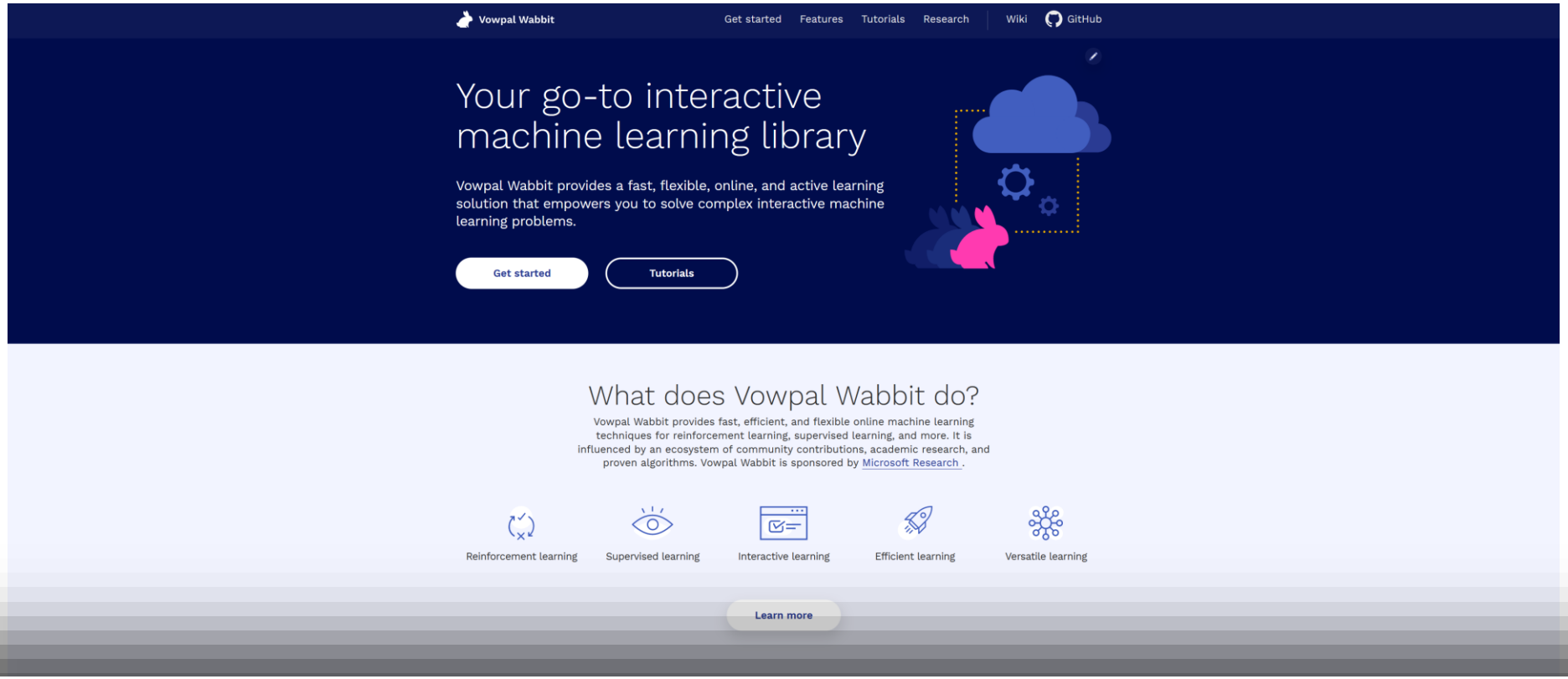
# Increased Investment

- More people working on VW
- We are focusing on improving and modernizing VW with stability and performance as priority
- Moved to VowpalWabbit GitHub organization
- New website
  - All new tutorial content

# Moved to the VowpalWabbit organization



# VowpalWabbit.org




tutorials, research, installation, help

# There’s never been a better time to get involved

# 8.7.0 – June 2019

- CMake build system (#1624)
- Coin betting (#1903)
- Contextual Memory Tree (#1799)
- Softmax learner for CB ADF (#1839)
- Cbify for CSOAA LDF datasets (#1681)
- Warm start for cbify (#1534)

# CMake – C++

- Easier to build from source

```
mkdir build
cd build
cmake ..
make vw-bin
```

- Easier to depend on

```
find_package(VowpalWabbit REQUIRED)
add_executable(my_exe main.cpp)
target_link_libraries(my_exe PRIVATE VowpalWabbit::vw)
```

# Coin Betting

- Efficient hyperparameter-free base learner
- Removes the need to tune hyperparameters
- https://arxiv.org/abs/1602.04128
  - Francesco Orabona, Dávid Pál
- Contributor - Francesco Orabona

# Contextual Memory Tree

- Learning memory store, with logarithmic time insert and retrieve
- Particularly useful for one shot learning
- https://arxiv.org/abs/1807.06473
  - Wen Sun, Alina Beygelzimer, Hal Daumé III, John Langford, Paul Mineiro
- Contributor – Wen Sun

# Softmax learner for CB ADF

- New CB type available to be used
  - `--cb_type sm`
- Softmax CB enables policy-gradient style reinforcement learning algorithms
- Contributor - Adith Swaminathan

# Cbify for CSOAA LDF datasets

- Cbify allows multiclass classification problems to be turned into CBs
- Add support for cbifying cost-sensitive classification datasets in LDF format
- Contributor - Alberto Bietti

# Warm Starting Contextual Bandits

- Ability to learn from set of supervised examples before transitioning to online learning
- Reduces regret when in online learning stage
- Wiki - https://github.com/VowpalWabbit/vowpal_wabbit/wiki/Warm-starting-contextual-bandits
- https://arxiv.org/abs/1901.00301
  - Chicheng Zhang, Alekh Agarwal, Hal Daumé III, John Langford, Sahand N Negahban
- Contributor – Chicheng Zhang

# 8.8.0 – December 2019

- Improvements to Python (#1928)
- Conditional Contextual Bandit (#1816) (#1995) (#2078) (#2141)
- CMake target improvements (#2172) (#2135)
- Internal refactoring to get ready for new changes

# Conditional Contextual Bandit

- Ability to fill multiple slots with actions
- Extension over contextual bandits
- Greater diversity and learning rate
- Natural representation for many personalization problems
- Wiki - https://github.com/VowpalWabbit/vowpal_wabbit/wiki/Conditional-Contextual-Bandit
- Contributors – Jack Gerrits, Yann Stadnicki

# What’s in the pipeline?

- C API improvements
- Bindings
- Slates multi-dimensional optimization algorithm
- Continuous actions support for contextual bandits

# Questions?

- Gitter - gitter.im/VowpalWabbit
- StackOverflow tag: vowpal-wabbit
- GitHub - github.com/VowpalWabbit/vowpal_wabbit
- Mailing list – groups.io/g/vowpalwabbit