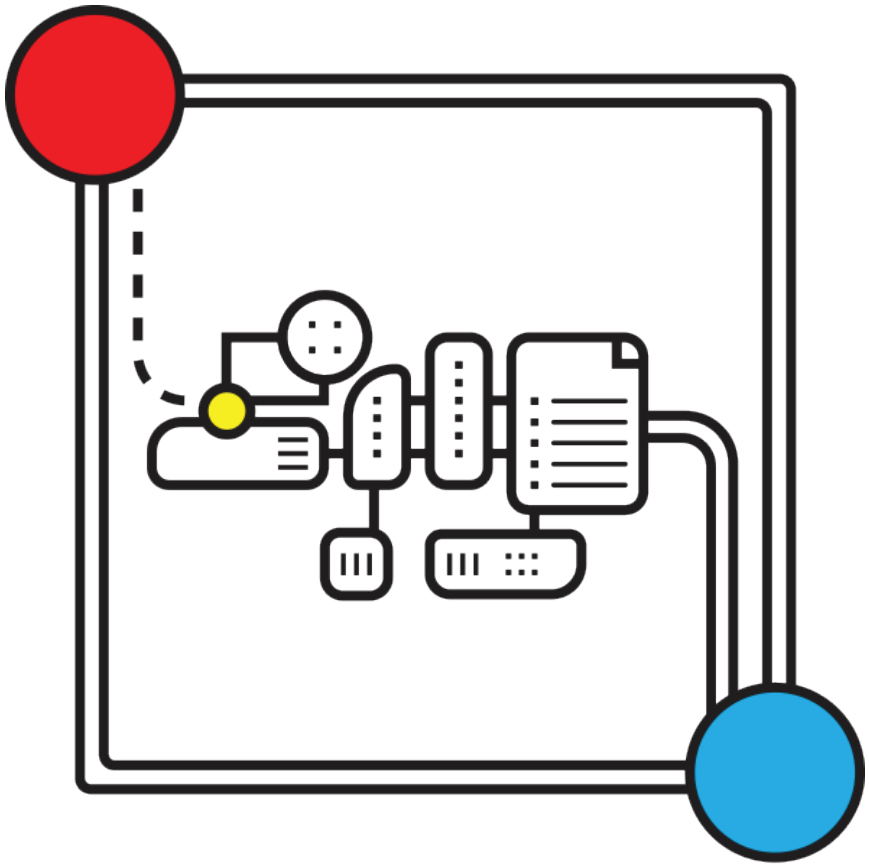


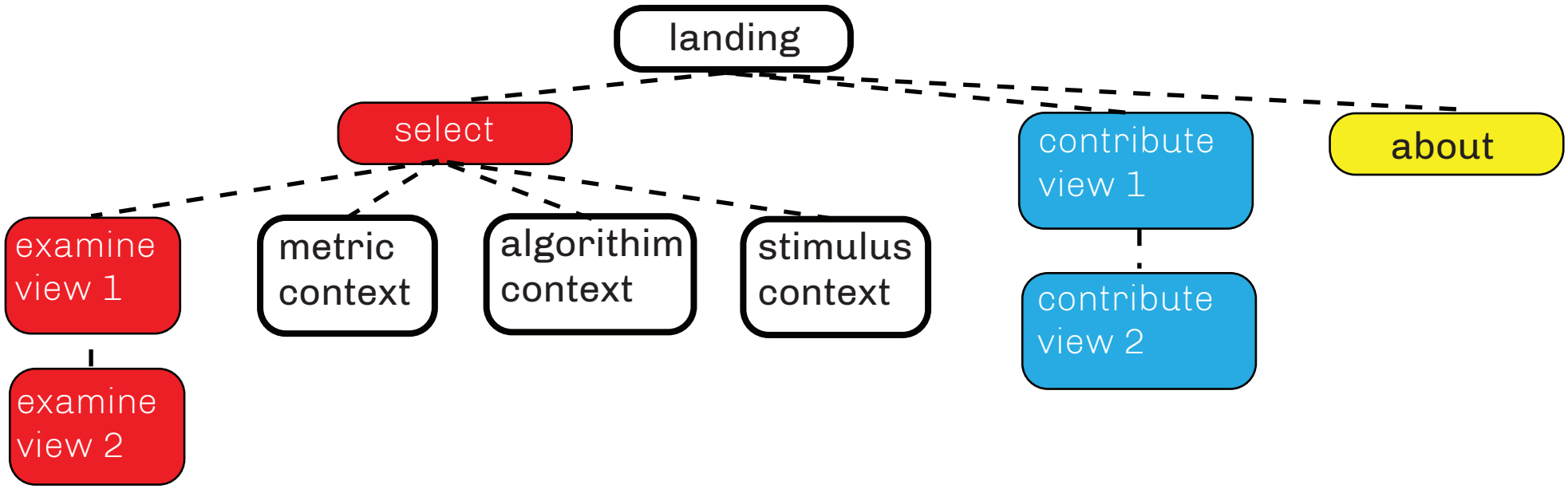
turing box functional specifications
V1.2 6.12.2018

introduction

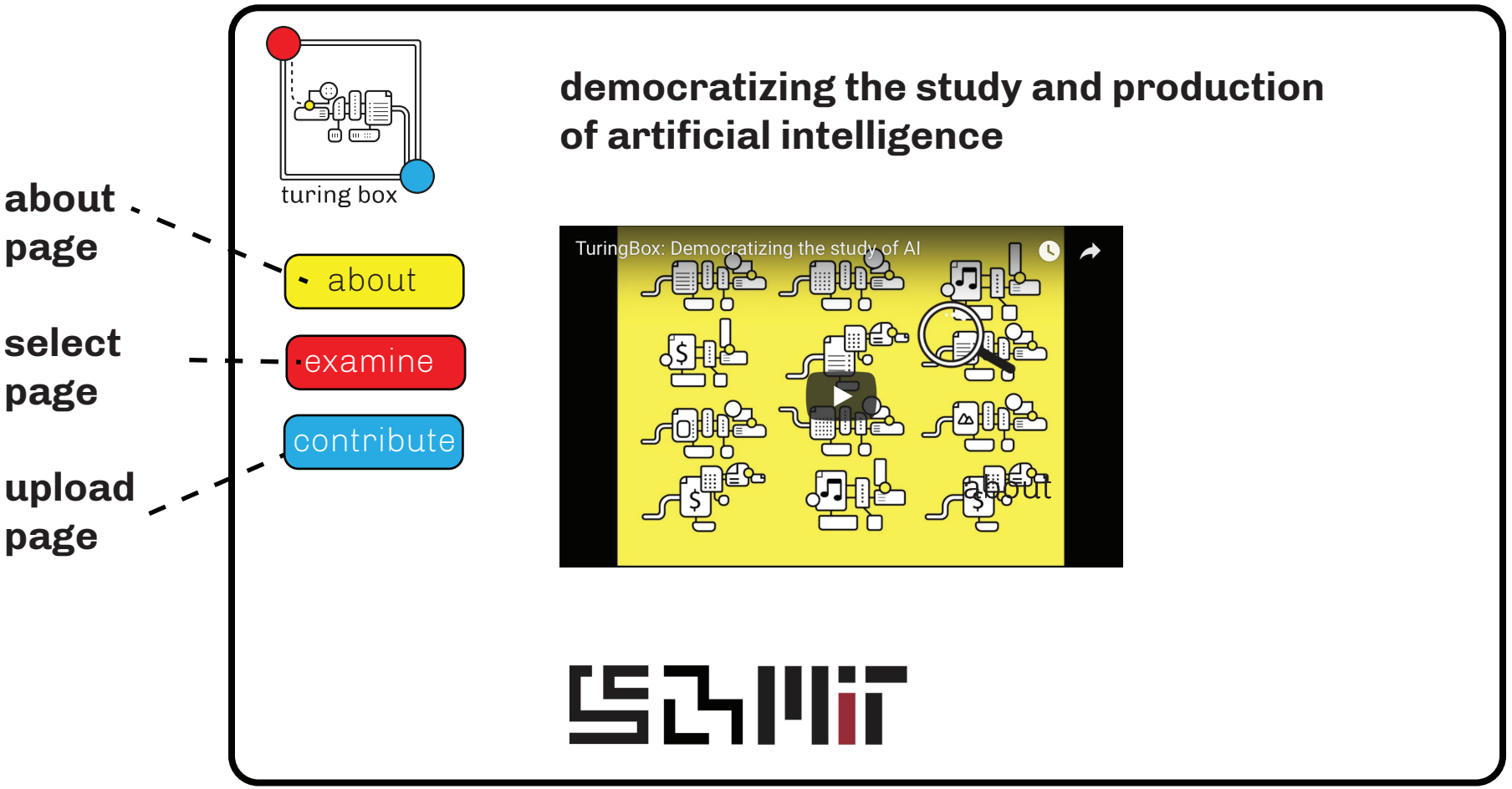
this document contains wireframes for the summer turing box development push. these designs are created to maximally articulate the following core values:

1. **conceptualization of algorithms as objects of scientific study**
2. **an impressive heterogenity of algorithms and datasets to study**
3. **a rigorous and comprehensive audit experience that points towards generality and sociality.**

the following pages and interactions have been constructed to best articulate these ideas, while providing a dynamic web experience.

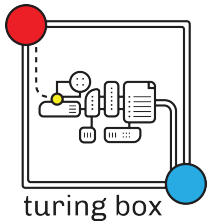


landing page



navigation buttons shrink on content pages

about page



turing box

about

examine

contribute

"Algorithms are being developed far faster than their impacts are being studied and understood. The Turing box...could help turn the tide. "
–Hal Hodson, The Economist

"The scientific study of machine behavior by those outside of Computer Science and Robotics provides new perspectives on important economic, social and political phenomena that machines influence. "
–Iyad Rahwan, MIT Media Lab, in The Nautilus

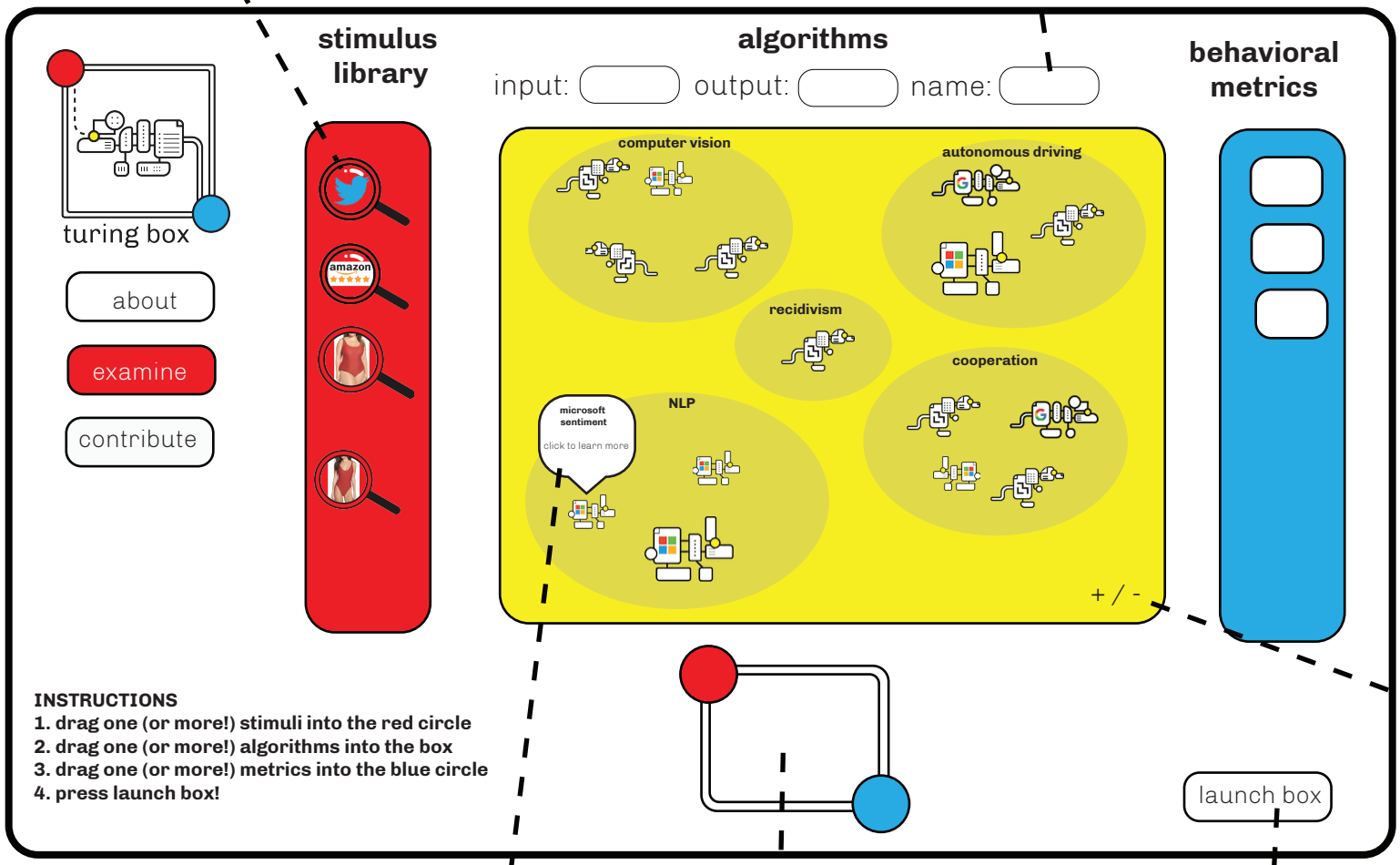
"Our diagnosis suggests that accelerating the scientific study of AI systems requires new incentives for academia and industry, mediated by new tools and institutions."
–Epstein and Payne et al.

more explanatory text and a faq

select page

client state holds input/output/name and dynamically displays stimuli, algorithms, and metrics according to filter

custom image or default graphic



algorithm map mode is zoomable

all simuli, algorithms and metrics have a description modal with a link to context page

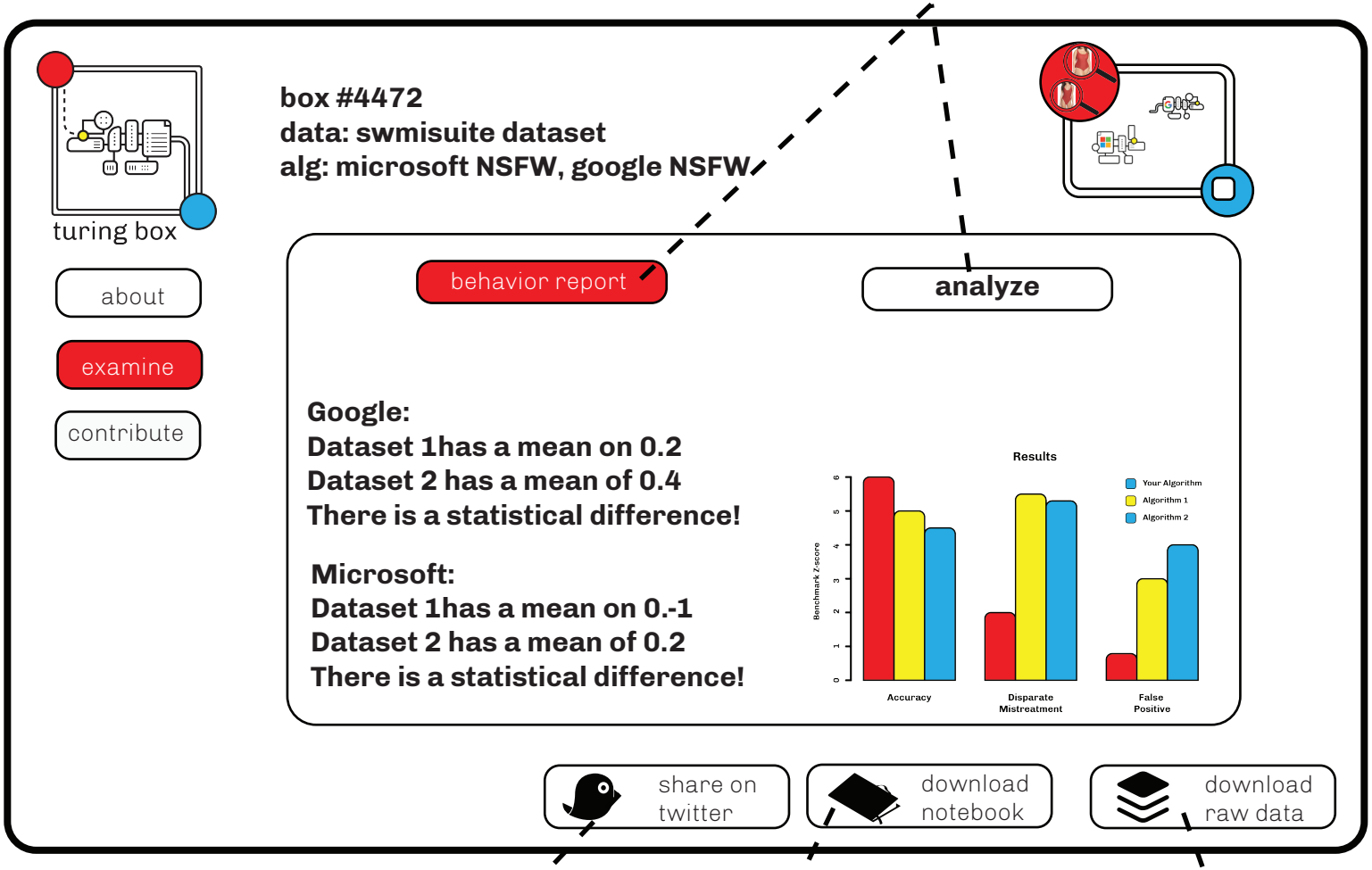
turing box is draggable space

launch box induces cool animation, spawns serverside task and takes you to examine page

examine page - view 1

box serialized by (dataset,alg) tuple so exmaine page simply renders pregenerated Compas Canonical data

toggle between views, default is report

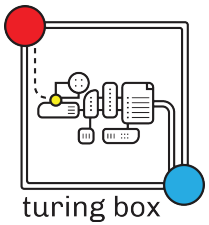


static turingbox url is sharable on social media

.pynb file for analysis

.csv of compas canonical data

examine page - view 2



turing box

about

examine

contribute

box #4472
data: twitter text corpus
alg: microsoft nlp

behavior report

analyze

```
In [1]: from bots import Bot
        import tasks
        from selection_tools import select_best_tweet

        bot_handles  = ["bots/Henry__Duboff","bots/HenryDuboff","bots/LiberalBot47","bots/Republic
        bots = []
        for bot_handle in bot_handles:
            bot = Bot(bot_handle, wait_on_rate_limit=True)
            bots.append(bot)

In [19]: #https://mashable.com/2014/04/04/celebrity-chefs-twitter/#orrbh_wsokgr
         mashable = ['altonbrown', 'Bordain', 'Rick_Bayless', 'Gachats','MarcusCooks','davidchang',
         #https://spoonuniversity.com/lifestyle/twitter-food-accounts-you-must-follow
         spoon_university = ['bonappetit','NatGeoFood','TwitterFood','Fooddotcom','newfoodeconomy',
         to_follow = mashable + spoon_university + ['FoodNetwork']

In [23]: for acct in to_follow:
             t = tasks.Follow(acct)
             for bot in bots:
                 bot.add_task(t,True)

         WARNING:root:The following task raised a Tweeperror: Follow Bot47Cooking, not scheduled
         Error : 403 Error message: Forbidden
         WARNING:root:Jose Campbell Dropped task: Follow Bot47Cooking, not scheduled

In [46]: tweets = bots[0].api.home_timeline()
         # Choose one with probability proportional to the tweet's score
         bests = []
         for i in range(0,100):
             best = select_best_tweet(tweets, punish_links=0)
             bests.append(best)

         hand_selected = [1004182595800850432, 1004604646405171202, 1004083045236953089, 1004000121
         bests = [1004373855266398208,1004350650329780225, 1004330719022178306, 1004366220274032641
         to_rt = hand_selected + bests

In [49]: for r in to_rt:
             t = tasks.Retweet(r)
             for bot in bots:
                 bot.add_task(t,True)

In [5]: import pandas as pd
        current_schedule = pd.read_csv("/Users/sive/GDrive/research/twitter-nudge/schedule_hyperdr

In [6]: day = 1
        for i,bot in enumerate(bots):
            targets = current_schedule[(current_schedule['day'] == day) & (current_schedule['bot']
            for j,target in enumerate(targets):
                try:
                    t = tasks.Follow(target)
                    bot.add_task(t,True)
                    print("bot {} target {} {}/250".format(i,target,j))
                except:
                    print("failed on bot {} target {} {}/250".format(i,target,j))
```

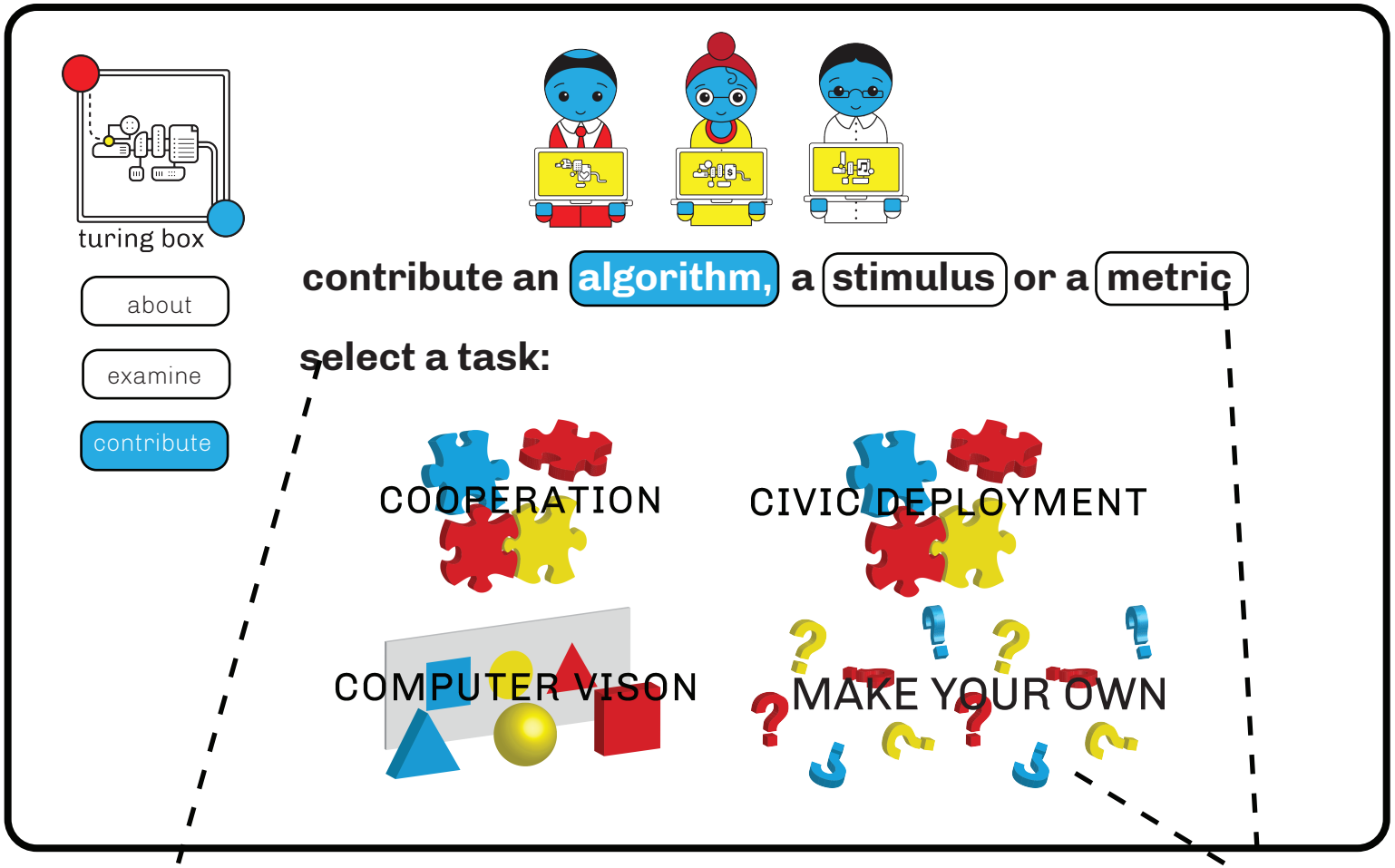
share on twitter

download notebook

download raw data

interactive juypter notebook iframe

contribute page - view 1

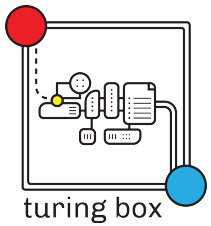


the task div is initially invisible. when algorithm or stimulus is clicked, the task div appears. it does not appear when metric is clicked

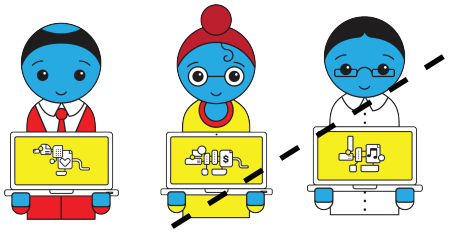
clicking a task or metric will take the user to contribute viewvv

contribute page - view 2

toggle between contribute views



turing box



about

examine

contribute

contribute an algorithm to the twitter sentiment task

how to contribute

on turing box, an algorithm is a .py file run from the commandline with arguments for the data.

for example, a valid algorithm for the twitter sentiment task be run as

```
python main.py twitter_dummy.csv
```

where main.py has a core() function that has as input a 28x28 .png file as input and as output a sentiment score (float) between [0.0,1.0]cv

click **here** to download the right dummy data for your task!

upload

name:

input: output:

description:

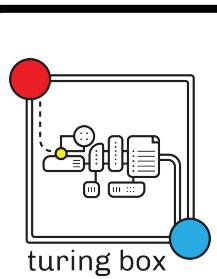
tags:

select file:

upload

algorithm context page

links to boxes the algorithm was used in



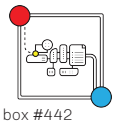
about

examine

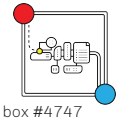
contribute

google cloud API

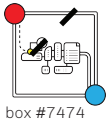
input: 28x28 .png image
output: Raciness score (float) [0.0,1.0]
description:
boxes:



box #442
swimsuite dataset
bias detected



box #4747
facial dataset
bias detected



box #7474
giraffe dataset
no bias detected

comments

this algorithm is used everywhere and has the potential for harm!

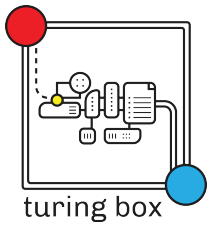
i found bias :(

all the studies that found bias in this are using imbalacned datasets! check out box #4747!

enter your messaged here

stimulus context page

block of text has wiki editting functionality



about

examine

contribute

swimsuite dataset - click to edit

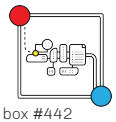
description: The Swimwear Model Dataset consists of over 200 matched pairs of images were scraped from online fashion swimwear catalogs such as Nordstrom, Dillards, Target, and more. Matched pairs consist of two images: one image of a standard-size model and one image of a plus-size model wearing the same swimsuit with similar lighting and poses.

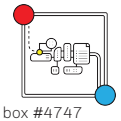
motivation for dataset creation: Why was the dataset created? (e.g., was there a specific task in mind? was there a specific gap that needed to be filled?) What (other) tasks could the dataset be used for? Has the dataset been used for any tasks already? If so, where are the results so others can compare (e.g., links to published papers)? Who funded the creation of the dataset?

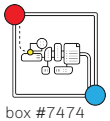
data preprocessing: What preprocessing/cleaning was done? Was the "raw" data saved in addition to the preprocessed/cleaned data? Is the preprocessing software available? tDoes this dataset collection/ processing procedure achieve the motivation for creating the dataset stated in the first section of this datasheet? If not, what are the limitations?

legal + ethical considerations: If the dataset relates to people (e.g., their attributes) or was generated by people, were they informed about the data collection? If it relates to people, were they told what the dataset would be used for and did they consent? If so, how? Were they provided with any mechanism to revoke their consent in the future or for certain uses? If it relates to people, could this dataset expose people to harm or legal action? (e.g., financial social or otherwise) What was done to mitigate or reduce the potential for harm?

boxes

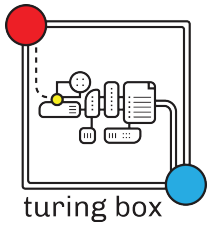

box #442
swimsuite dataset
bias detected


box #4747
facial dataset
bias detected


box #7474
giraffe dataset
no bias detected

comprehensive list of properties can be found at https://arxiv.org/pdf/1803.09010.pdf

metric context page



about

examine

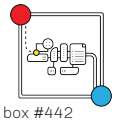
contribute

proportional parity - click to edit

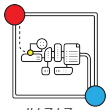
description: Ensure all protected groups are selected proportional to their percentage of the population. This criteria considers an attribute to have proportional parity if every group is represented proportionally to their share of the population. For example, if race with possible values of white, black, other being 50%, 30%, 20% of the population respectively) has proportional parity, it implies that all three races are represented in the same proportions (50%, 30%, 20%) in the selected set.

When does it matter? If your desired outcome is to intervene proportionally on people from all races, then you care about this criteria.

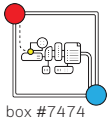
boxes



box #442
swimsuite dataset
bias detected



box #4747
facial dataset
bias detected



box #7474
giraffe dataset
no bias detected