FindingEmo: Supplementary Material

This document contains the required supplementary material for the paper entitled "FindingEmo: An Image Dataset for Emotion Recognition in the Wild".

Contents

- Changes From Previous Version
- Author Responsibility Statement
- Link to GitLab repository
- Link to Croissant metadata file
- Datasheet documentation
- Hosting, Licensing and Maintenance
- Reproducibility

Changes From Previous Version

A previous iteration of this work was submitted to ICML2024, and rejected as a result of the following main concerns.

1) The previous iteration of this work did not contain a "Limitations" section.

Addressed by: adding a "Limitations" section.

2) Some reviewers were concerned about the public part of our dataset having one annotation per image.

Addressed by: we have highlighted the fact that we follow existing work, in particular Emotic [1], in this regard. We also mention this limitation in our new "Limitations" section, and added a section to the Appendix (A.14 A Note on the Fuzziness of Emotion Recognition) that arguments why we believe this is not a major concern.

3) Concerns were raised with regard to the legal status of the dataset, in particular the question whether copyrighted material can be used without obtaining explicit approval from the copyright holders.

Addressed by: we have devoted a section of the Appendix ("A.2 Legal Compliance"), to refer to the appropriate legislation that indeed allows us to use copyrighted material for our research purposes.

4) The previous iteration of this work only contained results for CNN models, and some reviewers felt the lack of more recent ViT models to be too limiting.

Addressed by: we have included results for CLIP and DINOv2 models.

5) The previous iteration of this work phrased the Arousal and Valence prediction tasks as a classification problem instead of a regression problem, by binning the annotation values into three groups, and using these groups as classification targets. Some reviewers were not convinced by this approach, and would have rather seen these problems handled as regression problems.

Addressed by: redoing all experiments for Arousal and Valence prediction as regression problems. (Our GitLab repository contains all code necessary to obtain the old classification as well as the new regression results.)

[1] R. Kosti, J. M. Alvarez, A. Recasens, and A. Lapedriza, "Context based emotion recognition using emotic dataset," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, arXiv:2003.13401 [cs]

Author Responsibility Statement

The Author Responsibility Statement is included in the Appendix section of our paper, under A.15. What follows is a verbatim copy.

```
We, the authors, confirm that we bear all responsibility in case of any violation of rights during the
collection of the data or other work, and that we will take appropriate action if and when needed,
e.g., to remove data with such issues. We also confirm the licenses provided with the data and code
associated with this work: an MIT license for all code; a CC BY-NC-SA 4.0 license for the dataset
(concretely, the list of URLs and the annotations).

In particular, and as clearly and explicitly stated on our repository (under "Legal Compliance and
Privacy"), we invite any rightful copyright holders or persons depicted in any of the images that do
not want their work/likeness to be used within the context of this dataset to contact us, so that we can
remove that specific material from the dataset.
```

Link to GitLab repository

The GitLab repository containing all code and the dataset can be found at https://gitlab.com/EAVISE/lme/findingemo/.

Link to Croissant metadata file

The Croissant metadata file for our dataset is included in our repository at (raw content) https://gitlab.com/EAVISE/lme/findingemo/-/raw/main/croissant-findingemo.json?ref_type=heads. A copy of the file is also included with this supplementary material.

Datasheet documentation

This document includes dataset documentation for the FindingEmo dataset, following the "Datasheets for Datasets" approach by Gebru et al. (https://arxiv.org/abs/1803.09010).

It is also included in our repository at https://gitlab.com/EAVISE/lme/findingemo/-/tree/main/datasheet?ref_type=heads.

Motivation

1. **For what purpose was the dataset created?** *(Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.)*

The dataset was created to stimulate research in Emotion Recognition (ER) both from a Computer Science perspective (build computer models for ER) as from the perspective of Psychology and Neuropsychiatry (compare the workings of computer models to the human brain; use ER computer models to investigate ER within the human brain).

In particular, this dataset goes beyond existing image-based ER datasets that tend to focus on either the human face, or on a single individual. In contrast, we present a dataset that considers images *as a whole*, with each image presenting multiple individuals in various settings and forms of interaction.

1. **Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?**

The dataset was created by Laurent Mertens, while member of the KU Leuven [EAVISE](#) and [DTAI](#) research groups.

1. **Who funded the creation of the dataset?**

The creation of this dataset is part of the KU Leuven ID-N project "Computational Modeling of Social Cognition and associated Deficits by means of Artificial Neural Networks" with grant ID IDN/21/010.

1. **Any other comments?**

No.

Composition

1. **What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)?** (*Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions be- tween them; nodes and edges)? Please provide a description.*)

Each instance is (a URL to) an image accompanied by the annotation for that particular image. Each image depicts various people in various, naturalistic, social settings.

1. **How many instances are there in total (of each type, if appropriate)?**

We provide annotations for 25,869 images.

1. **Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?** (*If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable).*)

The publicly released dataset is a subset of larger set. In particular, we keep private a set of 1,525 with multiple annotations per image, which will allow us to organize dedicated future workshops (where this private set will be used as test set). This privately held set was also used to obtain inter-annotator agreement statistics, as reported in our [paper](#).

1. **What data does each instance consist of?** (*"Raw" data (e.g., unprocessed text or images) or features? In either case, please provide a description.*)

Each instance is essentially an unprocessed image accompanied by a row in a text-based CSV file representing the annotation for that image.

1. **Is there a label or target associated with each instance?** (*If so, please provide a description.*)

Yes, there are multiple targets associated with each image: emotion (following Plutchik's Wheel of Emotions); valence (named 'Negative/Positive' in the annotation interface, integer scale from -3 to +3); arousal (named 'Intensity' in the annotation interface) integer scale from 0 to 6); ambiguity (integer scale from 0 to 6); age group (of the people in the picture; multiple predefined labels possible); deciding factors (what made the annotator decide to go for this particular emotion?; multiple predefined labels possible) and optionally some specific descriptive tags.

An image of the annotation interface, with all annotated dimensions and associated labels/scales, can be seen below (displayed photo by David Shankbone; source: [Wikimedia](#). The interface also shows a 'Keep/reject image' choice, as annotators could opt to reject an image that did not

meet certain specified requirements. As all images in the dataset are 'Keep', this specific label was removed from the shared annotations.
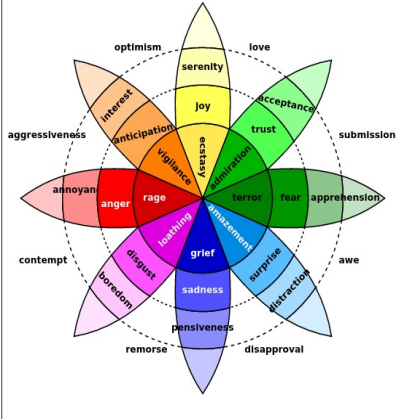
1. **Is any information missing from individual instances?** (*If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.*)

No.

1. **Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)?** (*If so, please describe how these relationships are made explicit.*)

Not applicable.

1. **Are there recommended data splits (e.g., training, development/validation, testing)?** (*If so, please provide a description of these splits, explaining the rationale behind them.*)

There are no recommended data splits. In the accompanying paper, we use a popular 80/20 train/test split. Importantly though, the distribution of the (emotion) labels is not uniform, so we do recommend that you take this into account when generating the splits. In particular, we advise to make sure that whatever split you chose is applied equally to each label, rather than randomly splitting the dataset.

1. **Are there any errors, sources of noise, or redundancies in the dataset?** (*If so, please provide a description.*)

Not that we are aware of.

1. **Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)?** (*If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (i.e., including the external resources as they existed at the time the dataset was created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.*)

The dataset links to external resources. In particular, we provide URLs to image files, rather than providing the image files themselves. Unfortunately, we cannot guarantee the persistence of these links. To mitigate this, we provide multiple URLs for as many images as possible.

With regard to restrictions, it is very important to note that these URLs point to images that are potentially copyrighted. In particular in the European Union, there is legislation (see Title II, Article 3 of the [InfSoc directive](#)) that allows the use of copyrighted materials for research and/or educational purposes *without* the need to obtain permission from the copyright holders beforehand. If you are not a member of a EU-based research institute, you will need to check with your local legislation whether you are allowed to use this dataset.

IN NO WAY CAN THIS DATASET BE USED FOR COMMERCIAL PURPOSES.

1. **Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)?** (*If so, please provide a description.*)

No.

1. **Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety?** (*If so, please describe why.*)

The dataset contains images depicting content that might be construed as sensitive to some consumers, such as violent scenes, scenes of bereavement, etc.

1. **Does the dataset identify any subpopulations (e.g., by age, gender)?** (*If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*)

Annotations contain labels for "age groups" of the people depicted in the images. An image can contain people from different age groups, hence annotators were allowed to indicate multiple labels. The (fixed) labels and their prevalence are: 'Children' (4588), 'Youth' (3460), 'Young Adults' (7541), 'Adults' (19457), 'Seniors' (5561), 'Undefined' (32). This last label indicates the image was not annotated with any label, which was a possibility early on in the data gathering process, before we made the annotating of all dimensions mandatory.

1. **Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset?** (*If so, please describe how.*)

Yes. All images are images depicting people, many of which are clearly identifiable. (We do not distribute the images directly.)

1. **Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)?** (*If so, please provide a description.*)

All images are images depicting people, which automatically reveals (or at least hints at) their race and gender. Nothing beyond the visible aspect is divulged or even known to us. (We do not distribute the images directly.)

1. **Any other comments?**

No.

Collection Process

1. **How was the data associated with each instance acquired?** (*Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.*)

The images were collected by means of a custom scraper, the code for which is included in this repository under duckduckgo_scraper.

The annotations were collected through Prolific. I.e., they were reported by human subjects through our custom annotation interface, a screenshot of which is included earlier in this document. Each annotator was asked to annotated 50 images, 5 of which, unbeknownst to them, were reference images (i.e., chosen and annotated by the main author, with the explicit goal of being as unambiguous as possible). Annotators were graded on these images, and if they graded too low, their annotations were discarded. For further details, please refer to our paper.

1. **What mechanisms or procedures were used to collect the data (e.g., hardware apparatuses or sensors, manual human curation, software programs, software APIs)?** (*How were these mechanisms or procedures validated?*)

See previous question.

1. **If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?**

As mentioned before, the (full) dataset is split into a publicly released part and a privately kept part. The dividing line between both is that the publicly released part has one annotation per image, while the privately kept part, which is much smaller, has multiple annotations per image.

1. **Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?**

For the image annotations, human annotators were recruited through Prolific, and were paid 10£ each for their efforts.

1. **Over what timeframe was the data collected?** Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.

The image pool from images were selected for annotation was gathered in the period from approximately 09/2021-06/2022. Collecting annotations through Prolific was done mostly in the period 09/2023-12/2023.

1. **Were any ethical review processes conducted (e.g., by an institutional review board)?** (*If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.*)

Yes, the data collection process was reviewed and approved by the KU Leuven Ethics Committee, in particular the Ethics Committee Research of University Hospitals Leuven, with dossier number S66479. Supporting documentation is not publicly available. Contact us for more information.

1. **Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (e.g., websites)?**

Images were collected from third party websites who were not notified of this (i.e., through scraping).

Annotations were collected directly from the human collaborators through Prolific.

1. **Were the individuals in question notified about the data collection?** (*If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.*)

As the annotations were collected through Prolific, a platform dedicated to high quality crowdsourcing, the participants were fully aware their answers were collected through our interface. No personal information was collected directly. The only personal information relating to the annotators we obtained was anonymized and provided by Prolific itself.

1. **Did the individuals in question consent to the collection and use of their data?** (*If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.*)

Participants had to agree to an Informed Consent clause before being allowed to proceed with the annotation task. A screenshot of the clause can be seen below.

Informed consent

Research title

Computational modelling of social cognition and associated deficits, through artificial neural networks

Name + contact details researchers

- Laurent Mertens (laurent.mertens@kuleuven.be)
- Prof. Hans Op de Beeck (hans.opdebeeck@kuleuven.be)

Research goal and methodology

The assessment of randomly shown images on a number of predetermined criteria. The goal of this research is to confirm unambiguousness in the assessment of images, as well as evaluate the intra-variability between different assessors. Finally, this research can potentially contribute to the selection of salient stimuli for a future fMRI study.

You will be shown a number of images, and you are asked to grade these on a number of criteria that were predetermined by us.

Experiment duration

We expect the experiment to take up to **1 hour** of a participant's time.

- I understand what is being expected of me during the experiment.

- I understand I will participate in the following tasks:
  ○ Assessment of a number of randomly chosen stimuli (photos), based on a number of criteria determined by us.

- There are no specific risks attached to the task.

- I understand I will be compensated in the following way for my participation:
  ○ Monetary remuneration through the Prolific platform.

- I understand my participation to this study is voluntary. I have the right to cancel my participation at any time, and do not have to provide a reason for doing so. No negative repercussions for my person can come from this.

- The results of this research can be used for scientific purposes and may be published. This includes the possibility of making the research data freely available on one or more open platforms. My name will not be published; anonimity and confidentiality of the data are preserved at every stage of the research.

- I known I can address any questions I have to:
  Professor dr. Hans Op de Beeck, hans.opdebeeck@kuleuven.be
  Laboratorium voor Biologische Psychologie, Tiensestraat 102 - bus 3714
  3000 Leuven, lokaal: 02.46 (tel: +32 16 32 60 39)

- For possible complaints or other concerns regarding the ethical aspects of this study, I can contact the Social and Societal Ethics Committee of KU Leuven: smec@kuleuven.be.

I have read and fully understood the above information and all my questions regarding this study were satisfactorily answered. I agree to partake in the study.

1. **If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses?** (*If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).*)

Yes, see previous question.

1. **Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted?** (*If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.*)

2. **Any other comments?**

No.

Preprocessing/cleaning/labeling

1. **Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)?** (*If so, please provide a description. If not, you may skip the remaining questions in this section.*)

Apart from the removal of the redundant "Keep/Reject" annotation (see Composition#5), no.

1. **Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)?** If so, please provide a link or other access point to the "raw" data.

N/A

1. **Is the software that was used to preprocess/clean/label the data available?** If so, please provide a link or other access point.

N/A

1. **Any other comments?**

No.

Uses

1. **Has the dataset been used for any tasks already?** (*If so, please provide a description.*)

Yes; see our paper for details, were we report on Emotion, Arousal and Valence prediction experiments using the dataset.

1. **Is there a repository that links to any or all papers or systems that use the dataset?** (*If so, please provide a link or other access point.*)

Not yet.

1. **What (other) tasks could the dataset be used for?**

The most straightforward application of the dataset is to train a computer model that can predict the dominant emotion/valence/arousal projected by a photo of people, focusing purely on the performance of the model.

Another application is again using the dataset to train such a computer model, but rather than focusing on the performance of the model, focusing on similarity to emotion processing within the human brain. This can be achieved by, e.g., comparing the output of the computer model with representations extracted from (f)MRI scans by means of an RSA procedure. This way, one can experiment with several existing and/or novel artificial neural network architectures, and see which one relates most, according to the RSA, to the human brain, which in turn can lead to insights into the functioning of the brain.

1. **Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?** (*For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other risks or harms (e.g., legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?*)

We did not perform an analysis of potential representation bias. In particular, the images are scraped from the internet at large, and hence, the risk exists that the data inherits the same biases. It could be possible that, e.g, images depicting confrontations between civilians and police predominantly involve civilians of a particular race more so than another, which might in turn lead to a model trained on this dataset associating people of this race more with negative emotions than positive ones.

1. **Are there tasks for which the dataset should not be used?** If so, please provide a description.

Yes. First and foremost, the dataset SHOULD NOT BE USED FOR ANY COMMERCIAL PURPOSE. Furthermore, we strongly advocate against the use of this dataset for any research with regards to surveillance/monitoring applications (e.g., crowd monitoring whereby either crowds or individuals are continuously monitored in real time in public spaces for their emotional state, with the potential goal of automatically detecting "potentially dangerous elements").

1. **Any other comments?**

No.

Distribution

1. **Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created?** (*If so, please provide a description.*)

Yes, the dataset is made publicly available to all.

1. **How will the dataset will be distributed (e.g., tarball on website, API, GitHub)?** (*Does the dataset have a digital object identifier (DOI)?*)

The dataset is distributed through the GitLab repository https://gitlab.com/EAVISE/lme/findingemo.

1. **When will the dataset be distributed?**

It was first put online on the 2nd of February, 2024.

1. **Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?** (*If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.*)

Yes, it is being distributed under a CC BY-NC-SA 4.0 license, which can be found in this repository, here.

1. **Have any third parties imposed IP-based or other restrictions on the data associated with the instances?** (*If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.*)

No.

1. **Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?** If so, please describe these

restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.*)

No.

1. **Any other comments?**

No.

Maintenance

1. **Who will be supporting/hosting/maintaining the dataset?**

The dataset will be hosted on GitLab, and maintained by Laurent Mertens. When L. M. departs from his position at KU Leuven, Prof. Joost Vennekens will become the point of contact.

1. **How can the owner/curator/manager of the dataset be contacted (e.g., email address)?**

By email at laurent.mertens@kuleuven.be and/or joost.vennekens@kuleuven.be.

1. **Is there an erratum?** If so, please provide a link or other access point.

No.

1. **Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)?** (*If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (e.g., mailing list, GitHub)?*)

We intend to expand the set of images for which we provide multiple URLs. These updates are expected to be sparse in frequency. Updates will be communicated on the GitLab frontpage (i.e., through the README.md file). No direct communication with dataset users is foreseen.

1. **If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were the individuals in question told that their data would be retained for a fixed period of time and then deleted)?** (*If so, please describe these limits and explain how they will be enforced.*)

There are no limits on the retention of data. This been said, rightful copyright holders or people depicted in the photos that do not wish their work and/or likeness to be used for the training of AI models can contact us to request that we remove the corresponding data from out dataset.

1. **Will older versions of the dataset continue to be supported/hosted/maintained?** (*If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.*)

The hosting of the dataset is expected to be persistent.

1. **If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?** (*If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/ distributing these contributions to dataset consumers? If so, please provide a description.*)

Yes, others can extend the dataset by using the exact same approach we have used! In particular, the custom annotation interface used by us to build the dataset is shared along with the dataset itself. Full details on the collection process are available in our paper.

1. **Any other comments?**

No.

Hosting, Licensing and Maintenance

Hosting

The dataset is hosted, together with the code, on our GitLab repository, https://gitlab.com/EAVISE/lme/findingemo/.

Licensing

We distinguish the code from the data. * Code: distributed under an MIT license * Date: distributed under a CC BY-NC-SA 4.0 license Both licenses are included in our repository.

Maintenance

(adapted from our datasheet documentation, as included in our GitLab repository) The dataset will be hosted on GitLab, and maintained by Laurent Mertens. When L. M. departs from his position in KU Leuven, Prof. Joost Vennekens will become the point of contact.

They can be contacted by email at laurent.mertens@kuleuven.be and/or joost.vennekens@kuleuven.be.

We intend to update the dataset by expanding the set of images for which we provide multiple URLs. These updates are expected to be sparse in frequency. Updates will be communicated on the GitLab frontpage (i.e., through the README.md file). No direct communication with dataset users is foreseen.

There are no limits on the retention of data. This been said, rightful copyright holders or people depicted in the photos that do not wish their work

and/or likeness to be used for the training of AI models can contact us to request that we remove the corresponding data from out dataset.

Others can extend the dataset by using the exact same approach we have used! In particular, the custom annotation interface used by us to build the dataset is shared along with the dataset itself. Full details on the collection process are available in our paper.

Reproducibility

Our GitLab repository contains instructions, described on the front page (README.md), on how to recreate the results reported in the paper. These can be found under the heading "How to obtain the results reported in the paper". In essence, all necessary Python scripts are provided. Users only need to adapt a config.py file to reflect their own configuration.