Civil War Twin: Exploring Ethical Challenges in Designing an Educational Face Recognition Application

Manisha Kusuma
Virginia Tech
Arlington, Virginia, USA
manishak@vt.edu

Vikram Mohanty
Virginia Tech
Arlington, Virginia, USA
vikrammohanty@vt.edu

Marx Wang
Virginia Tech
Blacksburg, Virginia, USA
boyuan@vt.edu

Kurt Luther
Virginia Tech
Arlington, Virginia, USA
kluther@vt.edu

ABSTRACT

Facial recognition systems pose numerous ethical challenges around privacy, racial and gender bias, and accuracy, yet little guidance is available for designers and developers. We explore solutions to these challenges in a three-phase design process to create Civil War Twin (CWT), an educational web-based application where users can discover their lookalikes from the American Civil War era (1861–65) while learning more about facial recognition and history. Through this design process, we operationalize a framework for AI literacy, consult with scholars of history, gender, and race, and evaluate CWT in feedback sessions with diverse prospective users. We iteratively formulate design goals to incorporate transparency, inclusivity, speculative design, and empathy into our application. We found that users' perceived learning about the strengths and limitations of facial recognition and Civil War history improved after using CWT, and that our design successfully met users' ethical standards. We also discuss how our ethical design process can be applied to future facial recognition applications.

CCS CONCEPTS

• **Applied computing** → **Arts and humanities**; • **Computing methodologies** → **Computer vision tasks**.

KEYWORDS

Human-AI interaction, facial recognition, digital history, ethical design, AI literacy, education

ACM Reference Format:
Manisha Kusuma, Vikram Mohanty, Marx Wang, and Kurt Luther. 2022. Civil War Twin: Exploring Ethical Challenges in Designing an Educational Face Recognition Application. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES'22), August 1–3, 2022, Oxford, United Kingdom.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3514094.3534141

AIES'22, August 1–3, 2022, Oxford, United Kingdom

ACM ISBN 978-1-4503-9247-1/22/08…$15.00
https://doi.org/10.1145/3514094.3534141

1 INTRODUCTION AND BACKGROUND

Facial recognition technology has established its presence in daily life, from its conceptualization in the 1960s as a military system, to its transformation into a highly commercialized product used in cities, smartphones, and airports [4, 17, 46]. Despite the widespread usage of these algorithms, researchers have drawn attention to their numerous flaws and underlying biases, especially against people of color, women, and non-binary people [9, 37, 43]. In recent years, these issues have provoked incidents and drawn public scrutiny, be it wrongful arrest of innocents [2, 21, 48], failing to verify rideshare drivers [7], or locking tenants out of their houses [15]. As the technology relies on a person's facial data, a form of personal identifiable information (PII), its usage has also raised concerns about data privacy and surveillance. For example, private companies collect photo databases without informed consent [16, 20, 22, 29] and law enforcement agencies use the technology to identify peaceful protesters [24, 25, 35]. Widespread coverage of these incidents and studies on facial recognition biases [9] have resulted in calls for legislation [14] and large companies refusing to sell the technology to police departments [18].

Despite its flaws and potential for misuse, facial recognition remains a powerful technology that can offer positive experiences to consenting users, be it for organizing personal photo collections, securing devices, or assisting low-vision communities [3, 54]. Since facial recognition's usage can be a "double-edged sword", it is important for the general public to understand both the risks and usefulness of the technology. Multiple surveys on facial recognition have shown the general public being divided on who should be allowed to use the technology and for what purposes, while sharing a unanimous concern about data privacy [23, 38, 44, 53]. However, these studies have also pointed out a major gap in the public's knowledge of facial recognition [45], and advocate for explainable interfaces with details around accuracy and data protection [38].

To raise awareness about face recognition and its limitations, several interactive AI projects have emerged in recent years. The "How Normal Am I" project draws attention to how AI can be used for judging faces, while highlighting the use of such algorithms in dating apps [41]. Wouters et al. created an interactive display called "Biometric Mirror" to provoke public reactions towards facial recognition being used for conducting psychometric analysis from a person's photograph [50]. Similarly, applications such as "Are You You?" [42] and "Emojify" [27] aim to inform the public about

automated emotion analysis from facial images. While these playful experiences are effective in highlighting undesirable outcomes of AI, they do not aim to simulate a realistic end-to-end user scenario that demonstrates how facial recognition algorithms are currently being deployed.

To address this gap, we developed Civil War Twin (CWT), an educational web application that aims to teach users about facial recognition and history through *learning by doing* [40]. The concept of CWT — users finding their lookalikes from the American Civil War era (1861–65) — was proposed to us by partners at the American Battlefield Trust, a non-profit historic preservation organization. This concept drew inspiration from Civil War Photo Sleuth (CWPS) [33], a web-based platform for identifying Civil War portraits using crowdsourcing and facial recognition; and Google's Art Selfie application that matches user selfies to similar-looking artwork [13]. In pursuing the educational goals of CWT, we encountered a number of underlying ethical challenges, including data privacy, transparency, gender and racial bias, and historical bias, that required careful design considerations.

In this paper, we address these underlying issues by proposing a three-phase, iterative ethical design process to develop CWT, where we first operationalized an *AI literacy* framework proposed by Long et al [28], followed by iteration and validation with academic experts and demographically representative potential users. We found that CWT's design successfully met the ethical standards of the users, and was effective in improving their perceived learning about Civil War history, and the strengths and limitations of facial recognition. We also discuss how our ethical design process can be applied to future facial recognition systems.

2 CIVIL WAR TWIN: CONCEPT AND DESIGN CHALLENGES

Drawing inspiration from the success of CWPS's face recognition-based identification workflow [33] and the Google Arts & Culture app [13], the American Battlefield Trust (ABT)[1] proposed a fun application for users to find and learn about their historical lookalikes from the Civil War era. This application, to be called *Civil War Twin (CWT)*, would help promote the visibility of their non-profit organization and contribute towards their mission of educating the public about the American Civil War.

The original concept of CWT was intended to be simple. A user would upload their selfie, and a facial recognition algorithm would search for similar-looking matches from a dataset of Civil War portraits, and display them in the form of downloadable artworks ("trading cards") for the user to share on social media or print at home. To reduce the application development overhead, we sought permission from the CWPS team to repurpose their existing facial recognition pipeline and photo database for CWT. Even though the user experience and implementation of CWT seemed straightforward at first, as we begun prototyping, we discovered that the original conceptualization presented several major ethical design challenges, which in turn led to opportunities for reframing the goals of the application.

[1] https://www.battlefields.org/

2.1 Design Challenges and Opportunities

2.1.1 ***Limitations of facial recognition algorithms.*** Black-box facial recognition systems provide little interpretability on how the input and output of a model are correlated [16]. In high-stakes scenarios like law enforcement, the "confidence score" determined by facial recognition algorithms has a direct effect on the safety and civil liberties of individuals in a society. Multiple studies have shown facial recognition has accuracy and bias problems, such as lower performance for dark-skinned or transgender faces [9, 43]. The black-box nature of such algorithmic models allows for racial and gender bias in real-time applications to go undetected. Raji and Buolamwini audited several corporate AI systems and found that despite recent improvements in classification systems, algorithmic bias is prevalent and continues to affect marginalized groups [37]

Without careful design considerations, these biases can negatively affect the user experience in applications such as CWT. Since the original concept for CWT would require a user's selfie as the only input to find their twins, it would inevitably rely on the algorithm's capabilities to automatically guess their ethnicity or gender, which is problematic for several reasons. First, the chances of inferring the wrong race or gender are unacceptably high [37, 43], and a wrong guess, even within a low-stakes scenario such as CWT, could offend a user's sense of identity or dignity and result in psychological harm [26, 31]. Second, even if the algorithm was accurate, it is hard to generalize what types of twins users will want to be matched with. For example, some users might prefer to see twins of a different gender, race, or army, while others might find these same matches offensive.

2.1.2 ***Biases in a historical database.*** Due to historical circumstances, ranging from a US Navy blockade to racial discrimination [10], some groups from the American Civil War era, especially Confederate soldiers, women, and people of color, have fewer surviving photos [12]. The CWPS database was seeded by public collections like the US Army's MOLLUS-MASS collection, which primarily contains portraits of white Union officers from northeastern states [33]. These historical biases were echoed in the composition of the original CWPS database and subsequently, the CWT database. This archival bias would lead to CWT users most likely getting matched to a white male Union soldier (88.3% of photos in CWT's database). One negative consequence is that the photos of other demographic groups, having a smaller reference pool, lack diversity. Users who choose to be matched to female twins, for example, will have lower-similarity matches, and many users will receive the same matches within these groups. Additionally, due to recordkeeping practices of the 1860s, gender on CWT is limited to the male–female binary, and soldiers' race classification is based on outdated historical legal frameworks such as the "one drop rule" or inferred from membership in racially segregated military units.

2.1.3 ***Concerns about user privacy and security.*** Multiple studies have reported user concerns about being exposed to facial recognition without informed consent [5]. Garvie et al. found that over 117 million American adults are present in law enforcement-based facial recognition databases, many without direct consent or probable cause [16]. Unless addressed explicitly, users of CWT might find themselves raising similar concerns about how their data is

being used, and who has access to their uploaded photos. With the rising use of facial recognition for biometrics, the subsequent faceprint created by the prototype during the facial detection step could potentially be misused and risk the security of users.

2.2 Ethical Design Approach

To address these challenges, we re-envisioned CWT as an ethically designed educational platform where users not only learn about Civil War microhistories and the contributions of various demographic groups, but also about the strengths and limitations of AI (i.e., facial recognition algorithms). Further, CWT also presents an opportunity to use the platform for educating users about the implications of privacy and surveillance in the context of facial recognition. This shift in focus from the original concept necessitated designing CWT as a platform that allows effective user *interaction* with the AI and fosters user *understanding* of the AI.

To achieve these goals, we adapted the design considerations proposed by Long et al. as part of the AI Literacy framework [28]. This framework, synthesized from a wide variety of interdisciplinary research, is aimed towards designing learner-centered AI, and consists of five umbrella themes — *What is AI? What can AI do? How does AI work? How should AI be used? How do people perceive AI?* — which are further composed of specific design guidelines.

For designing this new version of CWT, we incorporated these guidelines into a three-phase ethical design process where we 1) operationalized the AI literacy framework to design and build an initial prototype of CWT; 2) consulted experts in the fields of race, gender, and history to critically evaluate the initial designs; and 3) collected feedback from prospective users to iterate and validate the prototype. We executed the phases in this specific order to avoid "reinventing the wheel" by first leveraging existing knowledge and best practices, which was already documented and easily available, to tackle the broader design space. We then involved stakeholders to iteratively enhance the design, fill in oversights, and validate choices. This approach also helped manage the cost and time of the overall design process compared to alternative methods such as participatory design workshops.

3 PHASE I: OPERATIONALIZING THE AI LITERACY FRAMEWORK

Recent advancements in AI and machine learning have opened up exciting possibilities for enabling novel, beneficial forms of human-AI interaction. However, AI's complexity, unpredictability, and over-reliance on data pose numerous challenges for designing ethical and effective AI-infused applications [17]. Researchers have proposed multiple guidelines [1, 32, 36, 51] and AI fairness checklists [30] for designing AI-infused systems that address known issues around fairness, inclusivity, transparency, explainability, and privacy. Along similar lines, the AI literacy framework [28] proposes specific design considerations for building AI-based systems for fostering user understanding of the AI, which aligns with CWT's objective. In this phase, we operationalize this framework in the context of AI-based facial recognition to establish initial design goals for the CWT system.

3.1 Design Goals

DG 1: Promote transparency and explain how facial recognition is being used throughout the system. The AI literacy framework emphasizes the importance of transparency in all aspects of AI design, which includes eliminating black-boxed functionalities, comprehensive documentation about creator intentions, funding and data sources [28]. It also encourages incorporating explainable AI practices such as interactive demos and graphical visualizations to assist user understanding of the technology.

Adopting these ideas for designing CWT as a "white-box" educational facial recognition platform could help explain to users the inner workings of the underlying algorithm as the user goes through the process of finding their twins, including how faces are detected in an image, image search pools are constructed, and similar-looking images are retrieved.

DG 2: Allow user experimentation with the facial recognition algorithm in gradual steps. To help users understand a system's operations and how to interact with it, Long et al. proposed adopting design features that create the Sim-City effect [28] i.e., "a system that, through play, brings the player to an accurate understanding of the system's internal operations." The framework suggests allowing users the option to experiment and learn about different aspects of the system in *gradual steps* to avoid cognitive overload.

Breaking down CWT's user experience into multiple interactive steps corresponding to the different aspects of the twin search process could help users understand the underlying operations. For example, allowing users the option to filter the search pool could help them understand the role of reference databases in how a facial recognition algorithm operates.

DG 3: Make clear to the user the strengths and weaknesses of facial recognition. Long et al. argue that understanding the AI's strengths and weaknesses can help users better leverage the AI's capabilities [28]. Facial recognition algorithms tend to be less accurate for certain ethnic and gender groups, especially people of color, women, and non-binary people [9, 43]. Given that improving the algorithm's accuracy is beyond this project's scope, we can, however, enhance the educational experience of CWT users by highlighting the strengths and weaknesses of different aspects of the algorithm. For example, informing users about the algorithm failing to detect a face under poor lighting conditions could encourage them to upload clear, well-lit photos.

DG 4: Factor in how a user's identity might affect their experience with facial recognition. The AI literacy framework encourages designers to consider how a user's personal identity or cultural value might play a role in their learning experience with AI algorithms [28]. In the context of CWT, the underlying biases of both the facial recognition algorithm and the historical dataset can play a non-trivial role in affecting the experiences of users, especially underrepresented groups.

These issues point towards adopting design practices that give users more control over the algorithm by allowing for the customization of twin results through the selection of military service, gender, and ethnicity. We also choose not to use algorithmic-based

detection techniques to determine the ethnicity and gender of the user or the historical persons in our dataset.

In addition to algorithmic bias, there is also historical bias in the dataset of Civil War portraits we use. As noted above, many demographics lack substantial representation in the CWPS database. For the CWT system, we focus on educating users about these broad historical patterns of discrimination as well as microhistories of specific soldiers and civilians. Giving users a look into the individuals in our database and their stories could help highlight the contributions made by marginalized groups in the American Civil War. The use of graphical visualizations also provides an opportunity to educate users about the numerical distribution of photos and draw attention to the smaller demographic groups represented in the database.

DG 5: Collect and handle data responsibly. Long et al. also emphasize the importance of addressing key ethical issues surrounding AI, such as privacy [28]. In AI applications, privacy is based on providing notice and consent for how data is being used, and security focuses on mitigating user risk. With facial recognition systems, we need to protect the photos being uploaded and the subsequent faceprint created by the algorithm from being used in the public domain.

Given the rise in public concern over surveillance, data, and misuse [5], providing the user with information about how their data is used could help them make informed decisions about what data they are comfortable sharing. We can further establish user trust in the CWT system by limiting the amount of personal information collected to minimize security or privacy concerns, and by giving users control over how their data is shared and used.

3.2 System Description

Based on these design goals, we designed and developed a new, web-based version of Civil War Twin (see Figure 1). The website employs the educational technique of *learning by doing* so users can directly interact with AI to learn about facial recognition technology [40], meeting our *DG 2*. Through the process of discovering their look-alikes, users learn diverse microhistories from the perspective of persons who lived during the American Civil War *(DG 4)*.

Throughout the website, the *AI Text*, a side panel, provides more information about facial recognition technology *(DG 1)*. The top part of this panel has a "Behind the Scenes" section which explains in layperson's terms how the technology works, while the bottom part (i.e., "What Could Go Wrong?" section) describes potential shortcomings of the technology, historical records, or both. This panel allows for a more nuanced look at how facial recognition works and its limitations *(DG 3)*. Addressing *DG 4*, the website also provides historical text and links where users can learn more about the various demographic groups and individuals who lived in the US during the 1860s.

Users discovers their look-alikes ("twins") from the American Civil War era by following a four-step process: i) *Uploading a Photo*, ii) *Selecting Search Preferences*, iii) *Finding Matches*, and iv) *Discovering Twins*. We detail each step below.

3.2.1 Uploading a Photo. This page provides the user with context on how facial recognition detects facial features and explains how photo quality issues can affect the detection process *(DG 1)*.

The user begins by uploading their selfie, along with their name (optional) and email address, to the website (see B in Figure 1). The user is then asked to sign the IRB consent form outlining our study. Then, CWT uses Microsoft Face API to detect a face in the uploaded photo, displaying a bounding box around the face. If a face is not detected, the user is prompted to try again with another photo.

On this page, the user can also view CWT's Privacy Policy, which is accessible throughout the website, and describes what data is being collected and shared, and when it is deleted *(DG 5)*. The system deletes the user's personal information (email and photo) when the user exits the website, giving them control over their data. The user's photo and faceprint are not permanently stored on any web server (ours or Microsoft's) and are not used to train any facial recognition models.

3.2.2 Selecting Search Preferences. After uploading a photo, the user can specify search preferences (military service, gender, ethnicity) for the twins they would like to see (see C in Figure 1). The process is split across three pages as follows:

(1) *Military Service:* This page provides the user with information on how military service is determined for their potential twins. The user can then choose from the Union, Confederate, and/or Civilian categories to filter their pool of results. We use primary and secondary sources to identify and label the military affliction of the persons in our photo database.
(2) *Gender:* This page provides the user with information on how gender is determined for their potential twins and explains how gender bias affects facial recognition algorithms. The user can choose from the Man and/or Woman categories to filter their pool of potential twins. The gender of a twin is inferred from historical and medical records. The categories provided for the user reflect the historical dataset and thus, the practices of the era. For example, gender is limited to the male–female sex binary. Rather than obscuring these practices by mapping them onto modern-day labels, we employ seamful design [11] by presenting the historical context behind these categories and inviting the user to consider their own relationship to them.
(3) *Ethnicity:* This page provides the user with information on how ethnicity is determined for the potential twins and explains how racial bias affects facial recognition algorithm. The user can choose from the White, Black, Native American, Hispanic, and/or Asian categories to filter their pool of potential twins. Race in the context of the American Civil War era was determined based on a person's physical features and ancestry [34]. The five ethnic groups are reflective of the people represented in our photo database as soldiers' races were classified based on archaic legal systems and segregated military units of the era.

These search preferences help to mitigate the effects of algorithmic bias by providing an alternative for algorithmic-based detection techniques to determine the ethnicity and gender of the user or the historical persons in our database. We attempted to mitigate historical biases and increase inclusivity *(DG 4)* in several ways.

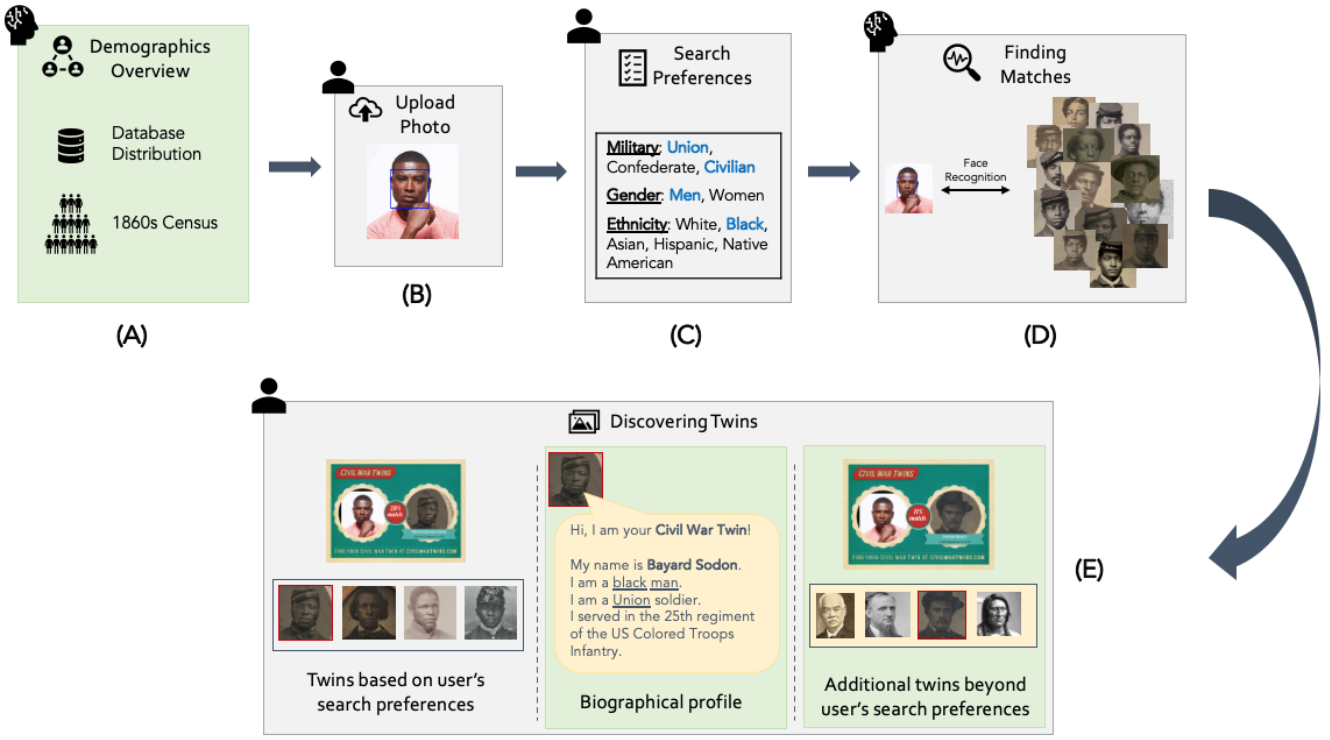

Figure 1: System workflow for the Civil War Twin website. The green sections were added in Phase II after receiving expert feedback. (A) Demographics Overview: The user can view graphical visualizations of the CWT database and its relation to the US population during the 1860s. (B) Uploading a Photo: The user can then upload a selfie of themselves to the website or select a set of stock photos to use. (C) Selecting Search Preferences: The user can specify search preferences (Military, Gender, Ethnicity) for the twins they would like to see. (D) Finding Matches: The user waits while the matching algorithm determines the top four similar-looking twins. (E) Discovering Twins: The user can see the four twins along with a trading card graphic. Biographical text is also added to provide information about each twin. The user can also view four additional twins outside their selected search preferences to learn more about the historical persons in the database.

First, we completed two targeted database enrichment projects to increase the number of photos of several underrepresented categories: African Americans, Asian, Native Americans, Hispanics, and women. Consequently, our collection of African American Union soldier portraits (128 photos), although comparatively small, is believed to be the largest digital collection of its kind in existence. Our database now has photos of 13,861 Union soldiers; 1,476 Confederate soldiers; 131 civilians; 15,358 men; 109 women; 15,272 whites; 17 Native Americans; 11 Asians; and 25 Hispanics.

Second, following *DG 1*, we created a real-time interactive visualization that shows how the user's preferences affect the search pool (see Figure 3 in Appendix). This visualization helps educate the user of our database construction as they are choosing their search preferences.

3.2.3 Finding Matches. Based on the military, gender, and ethnicity preferences the user selected, our algorithm determines the top four similar-looking twins from the database of reference photos. While the user waits a few seconds for the matching algorithm to find their twins, this interstitial "matching" page provides the user with information on how their face is being compared to their possible twins and how the confidence threshold affects twin results *(DG 1)*.

3.2.4 Discovering Twins. Once the twins are found, the user can view their four twin matches. This page, in accordance with *DG 1*, explains how the similarity score is determined for each twin (see Figure 4 in Appendix). CWT also emails the user a copy of their twin results.

The user can then view a trading card graphic (see E in Figure 1) for each set of twins. The user can download the trading card file for saving locally, printing, or sharing (see Figure 2 in Appendix). There is no direct way to share the trading card on social media as the system does not create persistent URLs as a privacy protection measure *(DG 5)*, but social media "share" buttons support manual sharing. The user can also learn more about their twins by clicking on their twin's CWPS profile links, which displays additional biographical and military records for that individual. Finally, the user has the option to contribute additional photos (e.g., from their

personal collections or public sources) to further enrich underrepresented categories in our database, or continue learning about AI through the additional links provided.

4 PHASE II: CONSULTING ACADEMIC EXPERTS

After implementing the redesigned prototype described in Phase I, we consulted three academic experts in Civil War History (E1), Gender Studies and Ethics (E2), and Race and AI (E3) to critique our design decisions. These experts, who were tenured faculty at our university, diverse in race and gender, and not previously familiar with CWT, helped us understand how the Phase I design engages with some of the sensitive topics around race, gender, history, and AI, as well as the broader societal implications of our proposed system. The experts' scholarly expertise and lived experiences helped (re)frame the design goals of our system and validate the design designs from Phase I. Based on their feedback, we iterated on our design and developed a high-fidelity prototype.

In separate one-on-one sessions, each expert was first introduced to the background, motivation, and goals of the CWT project, followed by a demo of the Phase I prototype, and a high-level overview of the Phase I findings and design choices. We then asked the experts about their general perceptions of the CWT prototype, along with a series of questions specific to their specialization . Overall, the experts found the "hook" of matching with one's Civil War twin, along with the concept of a trading card as a shareable proof of match, to be a fun and engaging experience, and appreciated the workflow of the application. The experts also favorably perceived the educational goals of the application and expressed that the Phase I design choices were effective in supporting these goals. At the same time, they also pointed out four focus areas for the system, outlined below, suggesting opportunities for improvement and updated design goals.

4.1 Expert Feedback and Revised Design Goals

DG 6: Provide more context for the creation and composition of the dataset, while highlighting the contributions of minority groups. All three experts found the real-time interactive visual chart on the search preferences page to be effective in conveying the demographic distribution of the database. Even though they appreciated the design efforts towards being transparent and acknowledged the challenges of historical bias, there were concerns about the dominant representation of white Union men over other groups. E1 said, *"The number and type of photographs of African Americans are going to be very different. So, I think just being really transparent is the right way to go"*. E3 pointed out the historical bias from a different lens by comparing the demographic distribution in the database to the actual population of the 1860s, stating that even if every woman in 1860s was photographed, the number would would still be disproportionately low compared to photos of men.

The experts also suggested other changes, such as using the more ethnically inclusive term "Black" as a category label instead of "African American." E2 stressed the importance of being upfront about the gender binary limitation of the dataset, while recognizing more fluid representations of gender in both the historical and modern eras. This suggestion aligns with the AI literacy framework's recommendation for contextualizing the dataset by providing more information about the dataset's composition and origins [28].

DG 7: Allow users to gain the educational experience without having to upload their photo. E1 raised concerns about the general sentiment in the media towards facial recognition technology and warned about possible public hesitancy towards CWT: *"There will be a substantial segment of the public who are not going to upload their photo regardless about what the site says about privacy"*. E1 emphasized the need to make the privacy policy easily accessible from every page and remind users about how their data is being used. Along similar lines, E2 raised concerns of users being wary of immediately uploading their personal photo as *"that might make them a little bit uncomfortable."* E2 further predicted that some people, out of general distrust towards large technology companies, might be cautious about sharing their photo with the Microsoft Face API. These concerns about privacy extend *DG 5*.

DG 8: Utilize empathy as a tool to understand history. All the experts identified a clear opportunity for CWT to use microhistories as a tool to tell individual stories from the Civil War. E1 encouraged us to use the CWT platform to foster connections: *"I do think there's value in just thinking about the connections between us as individuals and the people who lived through and fought the Civil War as individuals."* E3 challenged us further to not only form these connections but to build on the notion of empathy, enabling users to empathize with the various persons in our database and to understand their unique stories. As E3 stated, *"One conceptualization of empathy is that you want to stand in another person's shoes."* E2 added that CWT achieves a deeper level of learning that focuses on sensitivity and empathy which can be further developed through the use of microhistories.

When discussing the search preferences of military, gender, and ethnicity, experts believed that giving users more control over their potential matches was a justifiable approach. E1 talked about how users could be presented with unwanted twin results if there was no way of specifying preference: *"I think that's a really valuable feature giving users the choice, rather than you know, it is a potential minefield"*. However, E3 warned that giving users control of their search preferences can also lead to confirmation bias: *"It's quite dangerous to simply confirm people's existing biases."* E3 later expanded on an idea of "empathy twins," i.e., deliberately (with permission) showing twins results outside the user's selected preferences, which would aid in fostering connections and empathy as well as pushing back against stereotypes. This extends *DG 4* which was based on Long et al. [28]'s recommendation of considering a user's identity and values in designing interventions.

DG 9: Encourage users to speculate about the implications of facial recognition in the real world. E2 affirmed CWT's goal of educating users about the strengths and limitations of facial recognition. E2 said, *"It promotes public understanding of technology, and it does so not just by being a billboard or infomercial, but it gets them involved in using the technology"*. According to E2, the "learning by doing" aspect of our system could showcase the benefits of *"using the technology in order to help people understand the technology"*. While all the experts appreciated the informational panel about the AI Text, broken down into "Behind the Scenes" and "What

Could Go Wrong?" content alongside the interactive interface, they also raised concerns about the verbosity of the text. E3 suggested prompting users to question the technology. E2 elaborated on a similar notion: *"It's one thing to tell someone this technology has pros and cons. It's another thing for them to experience the pros and cons by getting a result."* Along similar lines, the AI literacy framework encourages imagining future applications of AI and their impact on the world [28].

4.2 Iterative System Description

Based on the feedback received by the experts and the new design goals we iterated on CWT by adding new features and modifying existing features.

4.2.1 AI Text & Speculative Questions. Inspired from speculative design approaches [49], we modified the AI Text throughout the website to include speculative questions around facial recognition *(DG 9)*. By employing simple elements of speculative design, users can begin to participate in the conversation about AI ethics and learn through an interactive process. The text also includes links to articles where the user can find other examples of facial recognition in the world and learn more about the technology.

4.2.2 Demographics Overview. We added a new visualization page prior to the initial photo uploading step (see Figure 1). This page provides the user with graphical representations of the CWT photo database across the different search categories (see Figure 5 in Appendix) and its relationship to the 1860s US Census. We were motivated to add this page to highlight the contribution of demographic groups *(DG 6)* and to be transparent about how our system is affected by historical bias. The page explains to the user how we collected the photographs in the dataset and the bias associated with them. The page also poses a speculative question — "Do you think your photo may be part of any such databases where facial recognition is being used?" — for the user to think critically about AI and existing public photo datasets.

4.2.3 Uploading a Photo. By continuing to be cognizant of how the system collects data *(DG 5)*, we added a new feature on this page that provides the user with a set of stock portrait photos to use instead of their own *(DG 7)*. If the user does not want to upload a personal photo, they can choose from a set of demographically diverse stock photos and continue the entire workflow to discover twins for the stock photo (see Figure 1).

4.2.4 Selecting Search Preferences. We updated the exemplars for each search preference type to present microhistories that reflect more diverse experiences *(DG 6)*. For example, for the "Gender" category, we included a wartime photo and biography of Albert Cashier, a Union soldier who was born female but lived as a man during and after the war.

4.2.5 Discovering Twins. For each twin, we added a new biographical, first-person prose on this page for the user to learn more about their twin's life while fostering natural connections *(DG 7)*. The prose contains information such as the gender, race, birthday, and military affiliation of the twin (see Figure 1). These basic identifiers help humanize the people in the photos to offer a point of introspection for the user. We also added a new page for "empathy twins", where the user can see additional twins beyond their search preferences. This new page shows the user four additional twins from a combination of search preferences (see Figure 1). By giving users an opportunity to learn more about the historical people in the database with whom they have a connection via facial similarity, but are demographically outside the user's selected search preferences, the system aims to foster empathy and perhaps change perceptions about who fought and lived during the Civil War.

5 PHASE III: COLLECTING PROSPECTIVE USER FEEDBACK

After implementing the designs that resulted from Phase II, we recruited nine prospective end-users (i.e., direct stakeholders) to use the CWT application and provide feedback on their experience via interviews. We recruited a demographically diverse set of participants to understand the effectiveness of the system's learning goals and gauge public perception of the system's ethical issues and our attempted mitigations. This qualitative study to evaluate CWT helped capture how users' perceived learning about the strengths and limitations of facial recognition and Civil War history improved how our design successfully met users' ethical standards, and how users interacted with the AI technology.

5.1 Methods

5.1.1 Recruiting Participants. We recruited nine participants (See Table 1) from different cultural organizations at our university, as well as online forums on Facebook and Reddit. We attached an interest form to our recruiting emails and blurbs that was used to screen participants based on demographics (e.g. race, gender, age), knowledge of facial recognition technology, and interest in history. We aimed to recruit participants with a diverse range of ethnicities, genders, and ages who were interested in United States history — representing the ABT's target audience — and had limited knowledge of facial recognition technology. The selected participants completed a consent form and a demographics survey, with IRB approval. The participants were compensated with a $20 gift card.

Table 1: List of Participants for Phase III

#	Gender	Ethnicity	Age Range
P1	M	Hispanic, Latino, or Spanish origin; Black or African American	30-39
P2	M	Hispanic, Latino, or Spanish origin	30-39
P3	NB	Asian	18-29
P4	W	White	50-59
P5	M	Black or African American	30-39
P6	M	White	40-49
P7	M	White	60-69
P8	W	White	70-79
P9	M	Black or African American	18-29

5.1.2 Procedure. Once participants consented to the study, they were given access to the CWT website and asked to provide availability for a 60-minute remote study session. Participants were

encouraged to familiarize themselves with CWT prior to the interview session. They could optionally send a screen recording of their experience on the website as they walk through the process of uploading their photo (or selecting a sample photo), selecting their search preferences, and viewing their twin results. This screen recording was intended to help us observe how new users interacted with the website and assess any bugs or unexpected behavior.

The interview session was conducted via Zoom video conferencing. We recorded the participant's video and audio with their consent. We gave the (two) participants who did not send in a screen recording ahead of time 15 minutes at the beginning of the interview to use the website while we screen-recorded their experience. The interview questions were divided into three main themes that corresponded to recurring topics throughout the project: facial recognition, Civil War history, and ethics. We also asked more targeted questions about specific features such as the database charts and the trading card graphic.

5.1.3 Analysis. We fully transcribed the interview audio recordings and used MAXQDA, a qualitative analysis tool, to organize participant's feedback. We used inductive thematic analysis to categorize participant quotes based on the themes that emerged from the transcripts [8]. We iteratively grouped together existing themes to organize quotes related to our main research topics.

5.2 Findings

Based on our analysis, we synthesized three main themes from the user study of CWT: understanding of face recognition, Civil War microhistories, and ethical values.

5.2.1 Understanding of Face Recognition. Prior to using the website, most participants had a general but limited awareness of how facial recognition technology works and is being used in society. After using the CWT system, participants perceived learning more about how the technology works and is being used, along with its strengths and limitations. The two main ways participants perceived learning was through direct interaction and reading the AI Text in the application. Below, we look at how the participants used CWT to improve their understanding of facial recognition.

Users explored how different inputs can affect the results of the face recognition algorithm. Eight out of the nine participants tested the website by uploading their own photos and one participant instead used the sample photos. Some participants even tried the site multiple times with different pictures of themselves to see if their twins differed. A majority of participants played around with their search preferences, specifically gender and ethnicity, to see how well their picture would match a twin of a different identity. P6 stated, *"I thought it was a really neat website, to be able to kind of play around and see do I match anybody, but also just the facial recognition technology was kind of neat to play with."* A minority of participants had a more directed, less playful approach. For example, P5 believed that changing his search preferences enabled him to refine the search results to achieve better matches: *"Maybe I have this wrong, but choosing and narrowing down the search on my end makes it that you actually could get something more accurate, is that the idea?"*

Users used facial features to determine the similarity of their twins. Participants focused attention on one or two high-diagnostic facial features to justify their twin results. For example, P8 noticed that her main twin had *"one eye that was more droopy than the other, as in mine, and I saw that, that was the connection."* P3 stated how the algorithm *"was able to detect accurately, like, femininity in my face"*. Participants also used such physical characteristics to gauge the accuracy of the confidence score produced by the face recognition algorithm.

Users considered how contextual factors can affect the face recognition algorithm. Some participants experienced firsthand how limitations in the dataset led to the low facial resemblance with their twin(s). For example, P2 strategically looked at multiple facial features to determine similarity and recognized the limited amount of Hispanic photos in our dataset: *"Slight similarities with like the eyes, maybe, but not too much like the chin [...] I think it was like a 36% match it wasn't a perfect match, but then again, I don't think [any] one looks like me at that time."* P6 pointed out how the relatively small size of the dataset could affect the accuracy of twin results: *"I didn't think that those pictures look that much like me, now granted this is a database of like as you said 15,000 photos as opposed to 100 million."*

Other participants, drawing on the AI Text and speculative questions, made connections to broader societal issues. For example, P3 acknowledged the importance of the dataset when trying to determine the accuracy and how in different contexts such as in policing it could be misused. P3 said, *"With this specific set of data and images I would not be able to determine whether or not it is accurate. Like I wouldn't be using this in like police facial recognition or anything like that."* Along these lines, some participants considered how the app's design surfaced their own biases, which could have negative impacts in higher-stakes scenarios. P5 pointed out that the application was fun within the context of historical twins, but alluded to a possible case of confirmation bias while comparing his selfie to his twin's photo: *"...maybe my brain is also making these connections that don't exist because this website told me that it has a 36% match."*

Users speculated about how facial recognition works and is currently being used. Participants referred to the AI Text for justifying the results of the facial recognition algorithm and how it works. When uploading a photo, P4 recognized, *"I was lucky because I had a good picture that really, I guess, must have just identified the characteristics pretty clearly [...] I mean it even talks about it on the left side of the screen, you know if your image is not that clear, your characteristics are not that clear, then there is going to be less of a chance that it's going to be a good match."*

Many participants explored the links provided and spent time to answer the speculative questions posed. P5, for example, answered as he uploaded his photo, *"If I was trying to identify someone which facial features, what would I pay attention to? The nose and mouth right? And facial hair? I think those are the main ones right."* P7 explored the articles linked on the AI text after reading the section: *"I think it was on the left-hand side that talked about where can facial recognition go wrong — false identification and false imprisonment. I was just reading some of the links that the program provided right*

before I logged in here, those are interesting." After using the application, P4 speculated about future applications of facial recognition technology: *"It just made me kind of curious how this possibly could become more utilized in the future. Like what if somebody looks very much like you, and they are using face recognition for security purposes or to enter a building. How does that work?"* Furthermore, P8 and P5 suggested areas of the AI Text where they wanted more details, such as adding more information about what facial landmarks were being compared during the matching process.

5.2.2 Civil War Microhistories. Most participants recalled an instance in K-12 education of learning about the Civil War as their main introduction to the subject. Some participants went on to learn more about the war from local museums, landmarks, genealogy, and personal research. CWT complemented this prior exposure of participants by allowing them to learn about the perspectives of different cultural groups and individuals during the 1860s and to empathize with their experience.

Users learned about the experiences of different cultural groups during the Civil War era. The process of selecting search preferences was effective in teaching participants about the different demographic groups that participated in the war. A majority of participants were surprised to learn about the critical roles that women played in the war. P1 thought it was interesting *"to know that women served, I didn't know to what extent."* P8 said, *"It's surprising to see that there were Asians and Hispanics that participated in this, and I was unaware of that. So that kinda opened up my eyes to what was going on."* P6 also described how CWT's experiential approach of teaching history is different from traditional mediums such as classrooms, textbooks, or articles: *This I think, brings it to life more so, instead of just being a date and a place [. . .] So many kids these days, are like, 'Oh, history is boring,' but there's more to it. There are real people, there are real consequences, there are real actions that happened. I think having this [CWT] kind of brought that to life more.*

Users formed a connection to their twins while trying to learn more about their lives. Participants explored biographical profiles on CWT, CWPS profiles, and other external sources (e.g., search results from Wikipedia, Google, etc.) to learn about their twins. P2 did not expect to find other sources of information about his twin, stating, *"This guy [Shawn Moffitt] was someone that was known and he lived through the war, which is a highlight to see [. . .] I looked up the name and stuff and then he popped up on Wikipedia and I was like, wow, because I wasn't expecting that."* P6 expressed a genuine curiosity for his twin: *"I wanted to learn more about him and I kind of wanted to go into it a little bit more to see, oh, what's his history? Do we know what happened to him? Did he survive the battles? And you know, does he have a family?"* One participant P4 even felt a familial relation to her twin, stating, *"She could have been my sister, you know. She looked like that!"* Some participants also sought more biographical information about their twin(s) directly on the website. P8 requested additional sources, saying, *"It would be nice to have some more information, you know, try these sites or, we got most of our photographs from here, try these sites."*.

Users wanted to discover new people from our database. Most participants, while using the website for the first time, did not select any search preferences to exclude or tailor results. Some felt it was the best way to test the system, while others did not want to restrict the facial recognition algorithm. P7 was comfortable about being matched with any results, due to his prior use of genealogy technologies:

> *I'm a little familiar with ancestry.com and their DNA database matches your DNA with your cousins. And I see all ethnicities and, of course, all sexes there, so I was accustomed to the fact that I could come up as a Hispanic or a person of color or a black person.*

Most participants, like P4, mentioned that they *"wanted to be open to anything"* out of curiosity. P6 enjoyed looking at additional twins outside of his search preferences even if the twins did not seem like an accurate match:

> *[The additional twins] gave me females, gave me people of color, gave me other things so, even though I may or may not match with someone else like that, it did give me a chance to also see what else there was and to really kind of bring home the point that there's more than just, you know, rich white men, you know there's more to it with Native Americans being involved.*

5.2.3 Ethical Values. Participants shared that they felt comfortable and safe using the platform, while recognizing the efforts taken towards creating an inclusive experience and being transparent about how personal information was being used.

Users felt the platform made successful efforts to be inclusive. P3, who identified as Asian and non-binary, appreciated the challenges of making CWT an inclusive experience: *"I'd say it's very difficult for there to be inclusion with any like system because the recording of historical information has been so white male-centered. [. . .] I think you've done the best that you could have done with this specific circumstance and set of data. So I thought it's pretty cool."* They were also aware of the limited number of gender and racially diverse people photographed during the 1860s: *"I don't think there's really like any data of gender diverse people. So, it's not necessarily something that you can accommodate for and you have only so much, like, racial data."*

Users felt the platform prioritized their privacy. Many participants expressed trust in the application because of the ubiquity of the privacy policy on the website. Most participants noticed the link to the "privacy policy" page, yet they never directly clicked to access the link, though they found comfort in its accessibility. P4 said, *"I skimmed over the privacy policy page because I was under the assumption and trust of this educational institution."* All participants felt comfortable uploading their photo to the website. The one participant who did not did not have a selfie available on his computer. P8 pointed to the website's explanation about not storing photos or faceprints, saying, *"I didn't have any problem, and particularly, when it says that your photo will not be saved."* Similarly, participants liked that the trading cards of their twins were emailed to them, instead of being saved on the system, not only because of the privacy preservation, but also because they could easily share them with friends and family via email.

6 DISCUSSION

6.1 Ethically Framing Facial Recognition Applications

Facial recognition applications are generally susceptible to ethical challenges related to privacy, gender and racial bias, and accuracy. In this paper, we employed a three-phase ethical design process to iteratively address these challenges. The feedback at each phase from a variety of internal and external stakeholders helped inform the design decisions applied to the system. In Phase I, we designed the initial CWT prototype by operationalizing Long et al.'s AI literacy framework [28]. This phase not only allowed us to position CWT as a learner-centered AI application, but also address underlying ethical challenges around transparency, privacy, and inclusivity.

In Phase II, the experts not only validated key Phase I design goals and features, such as the search preferences and AI Text, but also provided feedback complementary to the AI literacy framework, which led to new features, such as our database visualizations and "empathy twins". By recruiting a demographically representative set of potential users in Phase III, we were able to understand their experience on CWT and whether the system was effective in teaching them about face recognition and Civil War history. This phase was also an opportunity to understand the user's ethical values and determine if those values were met in the system design. Based on these findings, the participants largely validated the design decisions made in Phase I and II, so there were no major design iterations in Phase III.

Phase I is consistent with existing research on synthesizing human-AI guidelines for designing specific AI applications [47], but our experience adapting these guidelines for an educational face recognition application illuminated some of the unique ethical challenges of this technology. Future facial recognition applications can similarly reflect on how transparency, fairness, inclusivity, and privacy play a role in their system. Expert feedback as in Phase II can provide areas of improvement and specialized design goals for features that might have been overlooked. Consulting with a race and gender studies expert during Phase II provided insight into the current discourse around inclusivity and the roles of race and gender in technology [19].

Through this design process, we noticed that the feedback across all three phases were largely complementary, building on existing themes of privacy, inclusivity, and transparency. For example, the goals of inclusivity propagated through each phase. In Phase I, we made targeted efforts to increase diversity in our database, followed by the addition of demographics charts highlighted the minority representation in our database in Phase II. Finally, in Phase III, participants recognized the system's efforts towards maintaining an inclusive database despite historical limitations. Further, this design approach showed the potential of the AI literacy framework to shape the design of facial recognition-based applications to be consistent with the ethical values of multiple stakeholders.

6.2 Designing Interactive Digital Humanities

All participants cited their first introduction to the Civil War was in a classroom setting. Often the approach standardized in the K-12 school system is to focus on events, historical figures, and places, taught via the "pipeline model" of lectures transmitting facts, rather than more active or constructivist learning experiences. With CWT, we wanted to present history in a more interactive medium. With visualizations, we were able to provide participants with an overview of key information about the 1860s US Census. By exploring these charts, participants were surprised to learn about various demographics and who participated in the war. Also, when selecting search preferences, users were exposed to photos, biographical profiles, and anecdotes of historical individuals representing diverse cultural perspectives and experiences. This work aligns with a multicultural curricular approach that presents the experiences and perspectives of multiple cultural groups instead of only one mainstream view of American history [6].

Given the unique nature of the CWT database containing thousands of profiles of people who lived during the 1860s, we leveraged the idea of microhistories to tell human stories. Participants were able to learn about the war through the people who lived through it. A majority of the participants did not select any search preferences and were open to being matched with people outside their own identity fostering a notion of empathy. Empathy can be conceptualized in two different ways: putting oneself in the other's shoes and building a shared perspective [39]. Participants learned through the biographical text about the lives of civilians and soldiers living in this critical period in American history. Through the images and text, similarities were found (e.g. geographic location, ethnicity, gender, family history) to help form a bond between the twin and the user. The application of microhistories helps establish empathy as it frames large historical events into lived experiences. This concept is similar to other projects exploring how empathy can be induced through the use of AI [52].

7 CONCLUSION

Civil War Twin is an educational web application where users can discover their lookalike from the American Civil War while learning about facial recognition and Civil War history. We presented a three-phase ethical design process that documented how we operationalized a framework for AI literacy, consulted with different academic experts (in history, gender, and race), and collected user feedback for validating our design choices. We found that our system met the ethical standards of users and provided them an opportunity to learn about the strengths and limitations of facial recognition technology and Civil War history. CWT's workflow allowed users to directly interact with facial recognition technology, while the supporting AI text encouraged them to speculate about the implications of facial recognition. Users further displayed a level of empathy for their twins and were keen to learn more about their experiences during the Civil War. Our work opens the doors for research on navigating ethical design challenges with facial recognition applications, while demonstrating how topics such as AI and history can be incorporated into an interactive educational experience.

ACKNOWLEDGMENTS

We wish to thank Deborah Tatar, Sylvester Johnson, Paul Quigley, David Thames, and our study participants. This research was supported by NSF IIS-1651969 and a VT ICTAS Junior Faculty Award.

REFERENCES

[1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.

[2] Elisha Anderson. 2020. Controversial Detroit facial recognition got him arrested for a crime he didn't commit. https://www.freep.com/story/news/local/michigan/detroit/2020/07/10/facial-recognition-detroit-michael-oliver-robert-williams/5392166002/

[3] Apple. [n.d.]. Use the People album in Photos on your iPhone, iPad, or iPod touch. https://support.apple.com/en-us/HT207103

[4] Apple. 2021. About Face ID advanced technology https://support.apple.com/en-us/HT208108.

[5] Brooke Auxier and Lee Rainie. 2019. Key takeaways on Americans' views about privacy, surveillance and data-sharing https://www.pewresearch.org/fact-tank/2019/11/15/key-takeaways-on-americans-views-about-privacy-surveillance-and-data-sharing/. *Pew Research Center* (2019).

[6] James A Banks. 1993. Approaches to multicultural curriculum reform. *Multicultural education: Issues and perspectives* 2 (1993), 195–214.

[7] Eloise Barry. 2021. Uber Drivers Say a "Racist" Algorithm Is Putting Them Out of Work. https://time.com/6104844/uber-facial-recognition-racist/

[8] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[9] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*. 77–91.

[10] Mike Cason. 2020. Archives agency acknowledges distorting racial history. https://www.al.com/news/2020/06/archives-department-acknowledges-role-in-distorting-alabamas-racial-history.html Section: News.

[11] Matthew Chalmers, Ian MacColl, and Marek Bell. 2003. Seamful design: Showing the seams in wearable computing. (2003).

[12] Ronald S. Coddington. 2008. *Faces of the Confederacy an album of Southern soldiers and their stories*. Johns Hopkins University Press.

[13] Google Arts & Culture. 2017. Art Selfie https://artsandculture.google.com/camera/selfie.

[14] EFF. 2019. Bans, bills and moratoria. https://www.eff.org/aboutface/bans-bills-and-moratoria

[15] Lola Fadulu. 2019. Facial Recognition Technology in Public Housing Prompts Backlash. *The New York Times* (Sep 2019). https://www.nytimes.com/2019/09/24/us/politics/facial-recognition-technology-housing.html

[16] Clare Garvie, Alvaro Bedoya, and Jonathan Frankle. 2016. The Perpetual Line-Up | Unregulated Police Face Recognition in America https://www.perpetuallineup.org/.

[17] Kelly A. Gates. 2011. *Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance*. NYU Press, New York.

[18] Jay Greene. 2020. Microsoft won't sell police its facial-recognition technology, following similar moves by Amazon and IBM. https://www.washingtonpost.com/technology/2020/06/11/microsoft-facial-recognition/

[19] David Hankerson, Andrea R Marshall, Jennifer Booker, Houda El Mimouni, Imani Walker, and Jennifer A Rode. 2016. Does technology have race?. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 473–486.

[20] Drew Harwell. [n.d.]. FBI, ICE find state driver's license photos are a gold mine for facial-recognition searches. *Washington Post* ([n. d.]). https://www.washingtonpost.com/technology/2019/07/07/fbi-ice-find-state-drivers-license-photos-are-gold-mine-facial-recognition-searches/

[21] Kashmir Hill. 2020. Another arrest, and jail time, due to a bad facial recognition match. https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html

[22] Kashmir Hill. 2020. The Secretive Company That Might End Privacy as We Know It. *The New York Times* (Jan 2020). https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html

[23] Ada Lovelace Institute. 2019. Beyond face value: public attitudes to facial recognition technology. https://www.adalovelaceinstitute.org/report/beyond-face-value-public-attitudes-to-facial-recognition-technology/

[24] Amnesty International. 2021. https://www.amnesty.org/en/latest/news/2021/04/russia-police-target-peaceful-protesters-identified-using-facial-recognition-technology/

[25] Marc Freeman Joanne Cavanaugh Simpson. 2021. South Florida Police quietly ran facial recognition scans to identify peaceful protestors. is that legal? https://www.sun-sentinel.com/local/broward/fl-ne-facial-recognition-protests-20210626-7sll5uuaqfbeba32rndlv3xwxi-htmlstory.html

[26] Os Keyes. 2018. The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proceedings of the ACM on human-computer interaction* 2, CSCW (2018), 1–22.

[27] Dovetail Labs. 2018. Emojify https://emojify.info/menu.

[28] Duri Long and Brian Magerko. 2020. What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–16.

[29] Kim Lyons. 2020. ICE just signed a contract with facial recognition company Clearview AI. https://www.theverge.com/2020/8/14/21368930/clearview-ai-ice-contract-privacy-immigration

[30] Michael A Madaio, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach. 2020. Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[31] Kevin A McLemore. 2015. Experiences with misgendering: Identity misclassification of transgender spectrum individuals. *Self and Identity* 14, 1 (2015), 51–74.

[32] Microsoft. [n.d.]. Responsible AI https://www.microsoft.com/en-us/ai/responsible-ai.

[33] Vikram Mohanty, David Thames, Sneha Mehta, and Kurt Luther. 2019. Photo sleuth: Combining human expertise and face recognition to identify historical portraits. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 547–557.

[34] Vikram Mohanty, David Thames, Sneha Mehta, and Kurt Luther. 2020. Photo Sleuth: Identifying Historical Portraits with Face Recognition and Crowdsourced Human Expertise. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 10, 4 (2020), 1–36.

[35] Paul Mozur. 2019. In Hong Kong protests, faces become weapons. https://www.nytimes.com/2019/07/26/technology/hong-kong-protests-facial-recognition-surveillance.html

[36] Google PAIR. 2021. People + AI Guidebook. https://pair.withgoogle.com/guidebook.

[37] Inioluwa Deborah Raji and Joy Buolamwini. 2019. Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial ai products. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. 429–435.

[38] Kay L Ritchie, Charlotte Cartledge, Bethany Growns, An Yan, Yuqing Wang, Kun Guo, Robin SS Kramer, Gary Edmond, Kristy A Martire, Mehera San Roque, et al. 2021. Public attitudes towards the use of automatic facial recognition technology in criminal justice systems around the world. *PloS one* 16, 10 (2021), e0258241.

[39] Pier Giuseppe Rossi and Laura Fedeli. 2015. Empathy, education and AI. *International Journal of Social Robotics* 7, 1 (2015), 103–109.

[40] Roger C Schank, Tamara R Berman, and Kimberli A Macpherson. 1999. Learning by doing. *Instructional-design theories and models: A new paradigm of instructional theory* 2, 2 (1999), 161–181.

[41] Tijmen Schep. 2020. How Normal Am I? https://www.hownormalami.eu/.

[42] Tijmen Schep. 2021. Are You You? https://www.areyouyou.eu/.

[43] Morgan Klaus Scheuerman, Jacob M Paul, and Jed R Brubaker. 2019. How computers see gender: An evaluation of gender classification in commercial facial analysis services. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–33.

[44] Sovantharith Seng, Mahdi Nasrullah Al-Ameen, and Matthew Wright. 2021. A first look into users' perceptions of facial recognition in the physical world. *Computers & Security* 105 (2021), 102227.

[45] Aaron Smith. 2019. More Than Half of U.S. Adults Trust Law Enforcement to Use Facial Recognition Responsibly https://www.pewresearch.org/internet/2019/09/05/more-than-half-of-u-s-adults-trust-law-enforcement-to-use-facial-recognition-responsibly/. *Pew Research Center Internet & Technology* (2019).

[46] Francesca Street. 2019. How facial recognition is taking over airports https://www.cnn.com/travel/article/airports-facial-recognition/index.html. *CNN travel* (2019).

[47] Hariharan Subramonyam, Colleen Seifert, and Eytan Adar. 2021. ProtoAI: Model-Informed Prototyping for AI-Powered Interfaces. In *26th International Conference on Intelligent User Interfaces*. 48–58.

[48] Robert Williams. 2021. I Did Nothing Wrong. I Was Arrested Anyway. https://www.aclu.org/news/privacy-technology/i-did-nothing-wrong-i-was-arrested-anyway/

[49] Richmond Y Wong and Vera Khovanskaya. 2018. Speculative design in HCI: from corporate imaginations to critical orientations. In *New Directions in Third Wave Human-Computer Interaction: Volume 2-Methodologies*. Springer, 175–202.

[50] Niels Wouters, Ryan Kelly, Eduardo Velloso, Katrin Wolf, Hasan Shahid Ferdous, Joshua Newn, Zaher Joukhadar, and Frank Vetere. 2019. Biometric mirror: Exploring ethical opinions towards facial analysis and automated decision-making. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 447–461.

[51] Austin P Wright, Zijie J Wang, Haekyu Park, Grace Guo, Fabian Sperrle, Mennatallah El-Assady, Alex Endert, Daniel Keim, and Duen Horng Chau. 2020. A Comparative Analysis of Industry Human-AI Interaction Guidelines. *arXiv preprint arXiv:2010.11761* (2020).

[52] Pinar Yanardag and Iyad Rahwan. 2017. Deep Empathy https://deepempathy.mit.edu/.

[53] Shikun Zhang, Yuanyuan Feng, and Norman Sadeh. 2021. Facial Recognition: Understanding Privacy Concerns and Attitudes Across Increasingly Diverse

Deployment Scenarios. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*. 243–262.

[54] Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2018. A face recognition application for people with visual impairments: Understanding use beyond the lab. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

A APPENDIX

Figure 2: An example of the trading card created for each twin result.

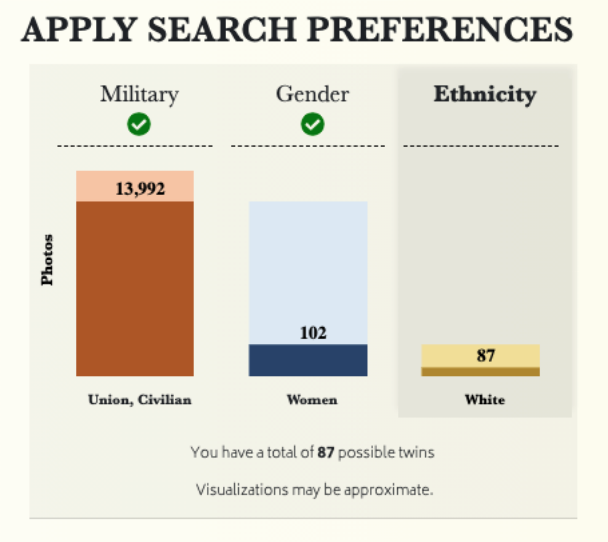

Figure 3: Graphical visualization of the database when the user selects the search preferences: Union, Civilian, Women and White.

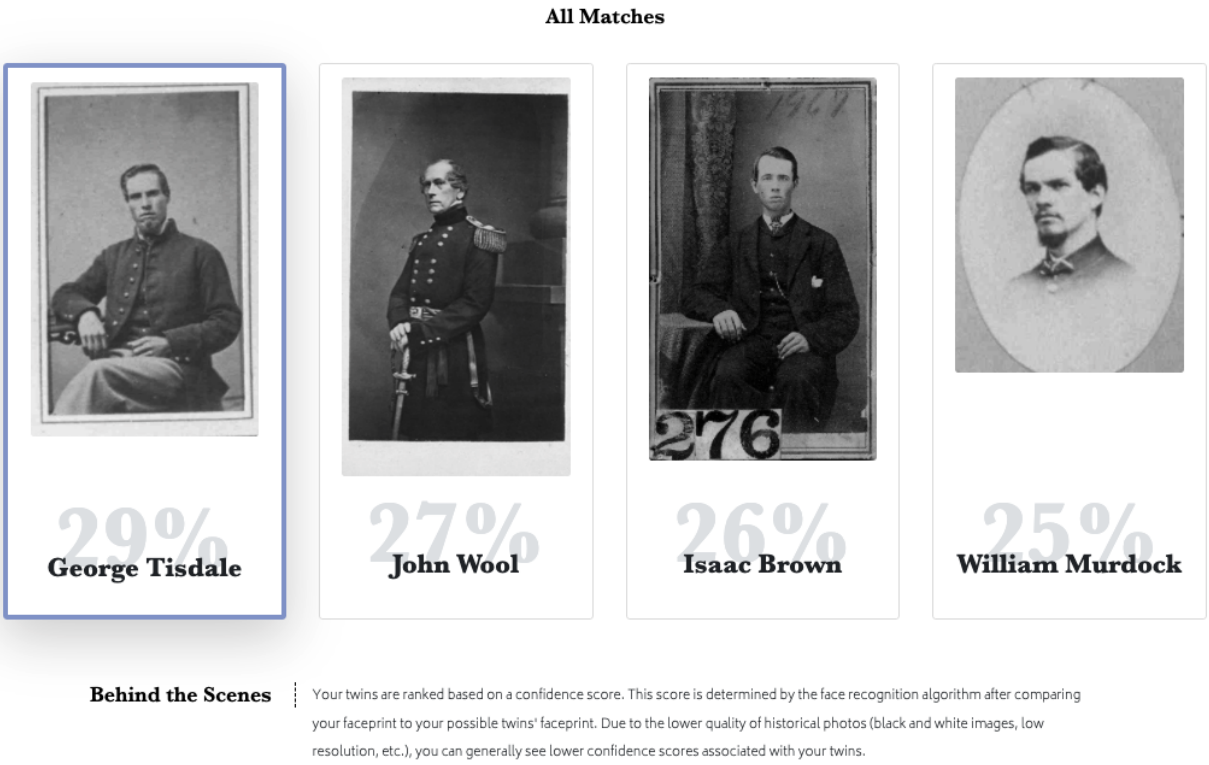

Figure 4: Screenshot of the "Discovering Twins" page with the four twin matches.

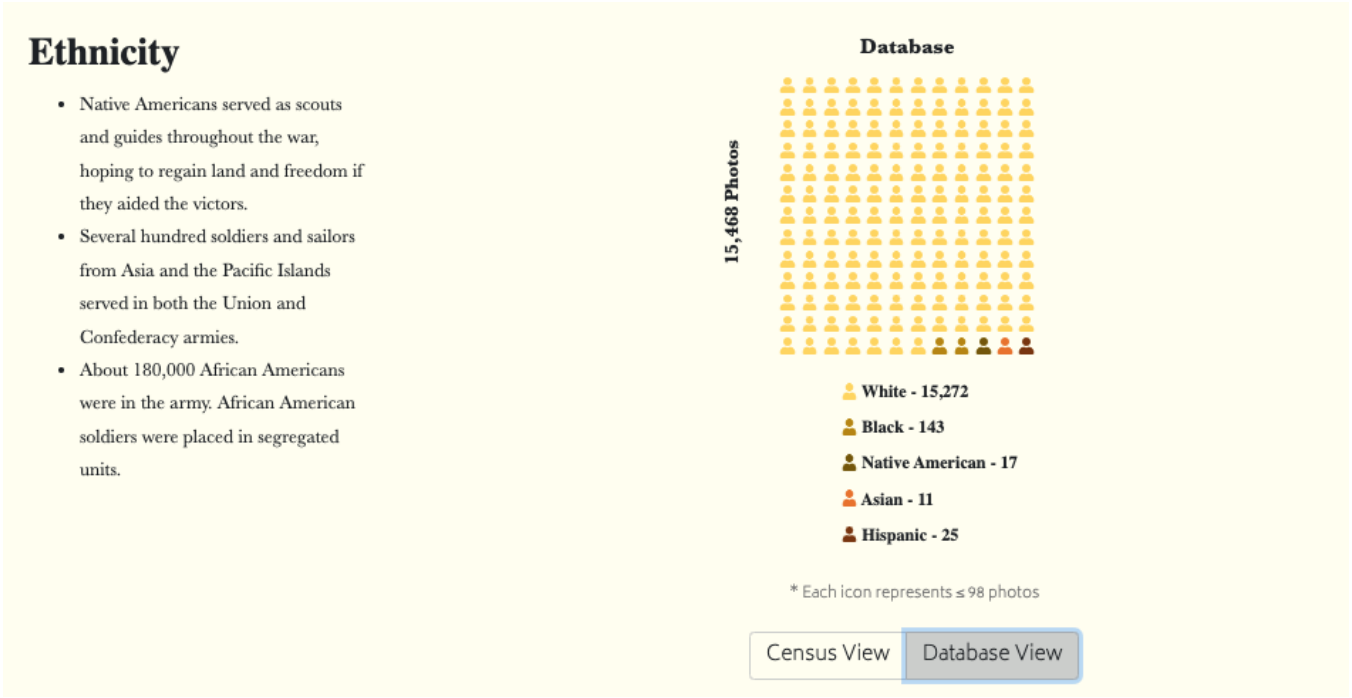

Figure 5: Screenshot of the "Demographics Overview" page with the Database visualizations of the five different ethnicities.

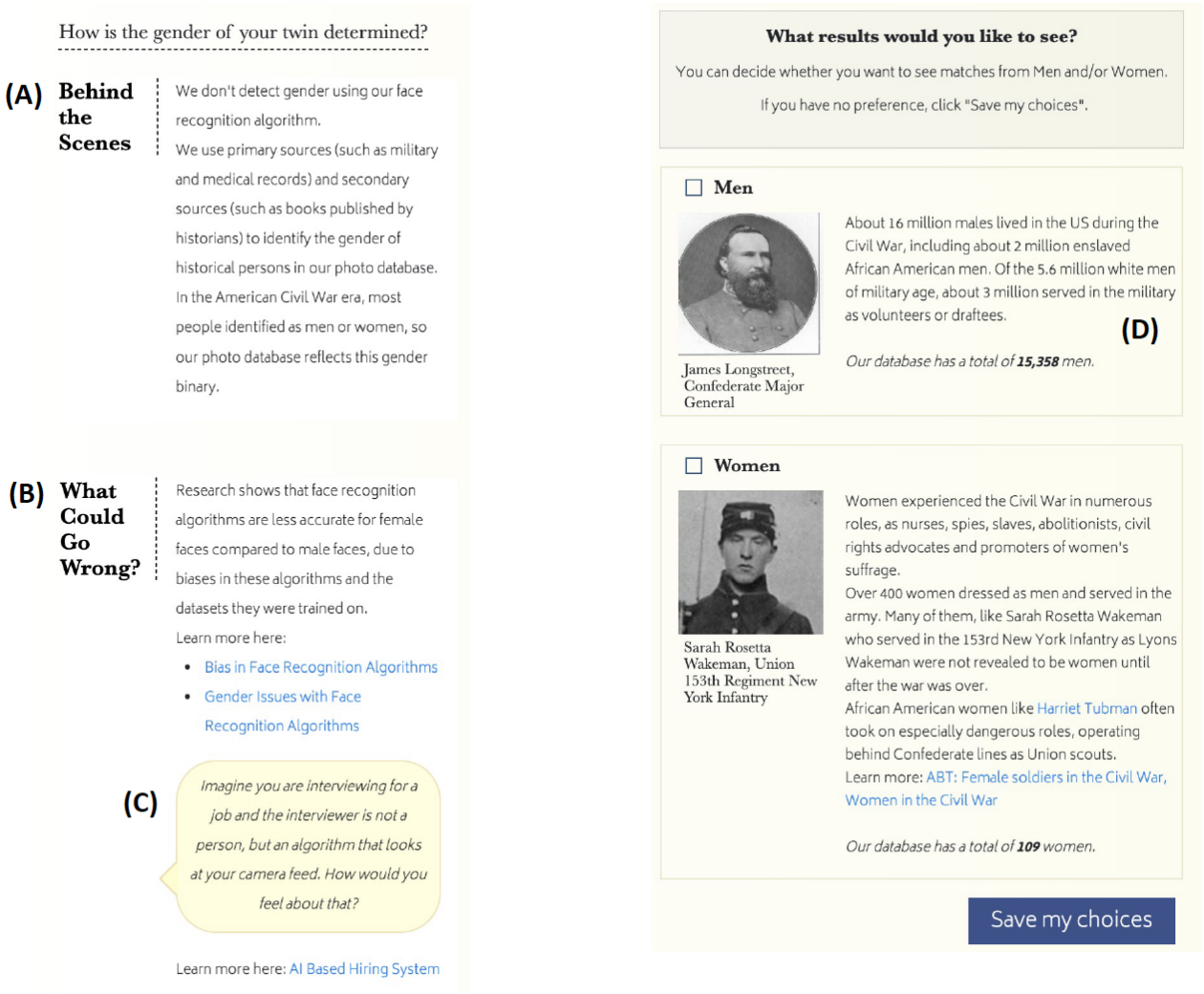

Figure 6: Textual content on the Civil War Twin website when the user is prompted to select their search preferences for gender. The left-hand panel is referred to as the AI text.
(A) Behind the Scenes: Part of the AI text that explains in layperson's terms how the technology works.
(B) What Could Go Wrong?: Part of the AI text that describes potential shortcomings of the technology (e.g., gender and racial bias) and/or historical records (e.g., historical bias). Includes links to further resources to learn more about AI.
(C) Speculative Question (Added in Phase II): Part of the AI text that prompts users to think and answers the posed question. Followed by a link to provide more context for the question.
(D) Historical Text: Part of every search preference, provides historical information and links to learn more about each identity during the 1860s.