1. Study Title

The GSA Equity Study on Remote Identity Proofing

2. Objectives of the study

The GSA Equity Study on Remote Identity Proofing aims to assess the impact of ethnicity, race, gender, income, and other demographic factors on multiple components of remote identity proofing, which is the process of verifying that a person is who they say they are.

GSA will test remote identity-proofing tools provided by multiple vendors that include both biometric checks using facial verification technology as well as non-biometric methods like mobile-device account ownership and credit history. The guidelines from National Institute of Standards and Technology's (NIST) Special Publication (SP) 800-63A for remote one-to-one identity proofing serve as a framework for the study.

GSA will release the study's results in a peer-reviewed publication. The publication will present a statistical analysis of failures and successes for the identity verification checks and explore the causes behind negative or inconclusive results. These results will help GSA understand the current technological barriers to equitable identity-proofing services for the public.

3. Background information

Why is the study being done?

GSA seeks to answer the question "Are commercially available identity verification systems equitable?" By combining the results of identity-proofing checks with participants' demographic information, GSA will be able to understand the magnitude and source of significant demographically-correlated disparities in the identity-proofing results.

3.1 Need

Federal cybersecurity requirements[1] require GSA to implement a single sign-on trusted identity platform for individuals accessing each public website of the agency that requires user authentication. The National Institutes of Standards and Technology's (NIST) Special Publication on Enrollment and Identity Proofing, requires such identity platforms to conduct a biometric comparison during the verification process to meet the requirements for Information Assurance Level 2 (IAL2)[2]. GSA is also mandated[3] by the Office of Management and Budget (OMB) to develop "...a roadmap for providing or updating GSA solutions and shared services

[1] 6 U.S.C. § 1523(b)(1)(D)
[2] *See* NIST SP 800-63A, para. 4.4.1.4 & Table 5-3.
[3] OMB Memo 19-17, page 11, [PDF, 13 pages]

that allow agencies to achieve the outcomes in OMB [Identity, Credential, and Access Management] policy and NIST standards and guidelines."

Insights from this information collection will enable GSA to make data driven decisions in the development of identity services needed for secure access to crucial government services. Without this data, GSA is at risk of deploying capabilities that fail to provide equitable access to government services for the American public.

3.2 Use

GSA will use the collected information to assess whether the currently-available identity-proofing technology can be used equitably to satisfy parts of NIST's SP 800-63 guidelines and thus inform policy and program development. GSA will achieve this by performing a statistical analysis of the multiple identity-proofing decisions made by the identity proofing systems included in the study platform. This analysis will be published in a peer-reviewed report for the benefit of the public and other federal agencies.

3.3 Academic Background

Most research in the field of biometric fairness has focused on the matching algorithm, i.e, the "set of instructions and rules for processing biometric samples" that performs the biometric comparison, i.e., "estimation, calculation or measurement of similarity or dissimilarity between a biometric probe(s) and a biometric reference(s)[4]". Buolamwini, et al. found that gender classification based on a single face image had a higher error rate for darker-skinned females with a high 34.7% error rate, compared to other groups (intersections of skin types and genders)[5]. Others found demographic differences in face recognition for some algorithms and systems[6,7]. This has led to extensive research on assessing variance across demographic groups for face recognition.[8,9,10]

---

[4] ISO/IEC 2382-37:2022(en) Information technology — Vocabulary — Part 37: Biometrics. https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:-37:ed-3:v1:en:term:37.08.03

[5] J. Buolamwini and T. Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," in *Proceedings of Machine Learning Research, Conference on Fairness, Accountability, and Transparency*, 2018, pp. 1–15.

[6] P. Grother, M. Ngan, and K. Hanaoka, "Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects," United States National Institute of Standards and Technology, Tech. Rep., 2019, NIST.IR 8280, https://doi.org/10.6028/NIST.IR.8280.

[7] C. M. Cook, J. J. Howard, Y. B. Sirotin, J. L. Tipton, and A. R. Vemury, "Demographic effects in facial recognition and their dependence on image acquisition: An evaluation of eleven commercial systems," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 1, no. 1, pp. 32–41, 2019.

[8] T. de Freitas Pereira and S. Marcel, "Fairness in Biometrics: A Figure of Merit to Assess Biometric Verification Systems," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 4, no. 1, pp. 19–29, 2022.

[9] J. J. Howard, E. J. Laird, Y. B. Sirotin, R. E. Rubin, J. L. Tipton, and A. R. Vemury, "Evaluating proposed fairness models for face recognition algorithms," *arXiv preprint arXiv:2203.05051*, 2022.

[10] K. S. Krishnapriya, V. Albiero, K. Vangara, M. C. King, and K. W. Bowyer, "Issues related to face recognition accuracy varying based on race and skin tone," *IEEE Transactions on Technology and Society*, vol. 1, no. 1, pp. 8–20, 2020.

According to the International Organization for Standardization (ISO) standard 19795-1[11], these types of studies would be considered a “technology evaluation” as opposed to a “scenario evaluation”. A scenario evaluation adds the full context of the use case and includes the end-to-end system which incorporates image-capture hardware and software, user experience, quality control, etc. The GSA study would be considered a scenario evaluation, as it is an end-to-end evaluation of the software.

Facial verification in the GSA study is not simply matching two images (as in a technology test), but includes the full end-to-end experience of the individual. By implementing an end-to-end scenario evaluation of the remote identity proofing solutions, GSA will be able to learn whether or not the legitimate users' experiences measured by performance metrics varies by demographic group.

The recent Department of Homeland Security Biometric Technology Rallies[12,13] explored scenario evaluations for one-to-many matching. In these studies, participants’ pictures were captured with a high resolution camera under ideal lighting to create a reference “gallery”. This gallery was then incorporated into multiple face recognition systems. Participants would then enter a “check-in/security gate” environment either by themselves or in groups while each of the systems under test would capture one best photo of the participant and try to recognize and match these images to the people from the reference gallery.

The proposed GSA study leverages the design and architecture of the DHS’s Rally with important distinctions:

1. The GSA study is not conducted in a laboratory setting and thus the data will feature a wide range of imaging conditions (ie. camera model, environmental lighting and setting, participant’s proficiency taking pictures, etc.);
2. The GSA study is recruiting a larger participant pool (4,000 versus DHS’s ~600); and
3. The GSA study expands on the identity proofing methods under test by incorporating consumer history checks that look for identity markers in physical address and financial records as well as phone account validation and device risk assessment products that are not usually tested.

GSA’s study is the first of its kind in the U.S.; the study will attempt to gather sufficient data to determine the severity of bias in remote identity-proofing scenarios under “real world” conditions. GSA will recruit two to five times more participants than similar studies; obtain representation from all parts of the United States; and ensure that participants are evenly distributed across demographics. To GSA’s knowledge, this is the largest public scenario evaluation of its kind.

---

[11] ISO/IEC 19795-1:2021(en) Information technology — Biometric performance testing and reporting — Part 1: Principles and framework.https://www.iso.org/standard/73515.html

[12] The 2021 Biometric Technology Rally (2021 Rally) at MdTF. https://mdtf.org/Rally2021/Results2021?Length=0. Last Accessed: April 6 2023.

[13] The 2022 Biometric Technology Rally (2022 Rally) at MdTF. https://mdtf.org/Rally2022/Results?Length=0. Last Accessed: April 6 2023.

3.4 Method

GSA will collect consent, demographic information, identity-proofing results, and exit survey feedback electronically. The demographics information provided by participants is used to determine participant eligibility based on the current status of the needed statistical sampling populations. Eligible participants will then be asked to participate in facial verification and other identity verification processes.

The identity-verification part of the study is accomplished through a GSA-developed system known as the Identity Verification API (IDVA). IDVA integrates multiple commercial products that have been vetted by GSA's Privacy and Security teams. All data will be stored in an encrypted GSA Google Drive that can only be accessed by a controlled group of GSA civilian and contractor personnel.

4. Recruitment methods

How subjects are identified and recruited.

4.1 Recruitment plan

GSA is taking a two-pronged approach by working with a recruitment agency as well as engaging with community organizations.

GSA will be working with a recruitment partner, Rekrewt to engage the general American public to participate in the study; engagement includes publishing social media advertisements and recruit participants across:

- Race/Ethnicity
- Gender
- Age
- Income
- Education.

Furthermore, GSA will work directly with various community organizations that work with these populations to leverage their mailing lists and other communication methods.

4.2 Pre-participation information collection

Screening for the study is done in two phases. First, GSA asks participants to become familiar with the study steps and the information being collected. If participants consent, GSA asks them for their name and email address. This information is used to send participants an email with a hyperlink to the demographics screening survey. This link is available for 24 hours.

The demographics screening survey asks participants to provide the following information from each participant.

1. Select Ethnicity:
    - Hispanic/Latino
    - Not Hispanic or Latino
2. Race
    - White/Caucasian
    - African American
    - Hispanic/Latinx
    - Asian, Hawaian, or/ Pacific Islander
    - American Indian or Alaskan Native
3. Gender
    - Male
    - Female
    - Another gender identity
4. Age (direct numerical input)
5. Household Income Range
    - $0- $24,999 / year
    - $25,000 - $75,999 / year
    - $76,000 - $99,999 / year
    - $100,000 and higher / year
6. Education Levels
    - Some high school or less
    - High school diploma or GED graduate
    - Some college or university, but no degree
    - Bachelor's degree
    - Graduate or professional degree (MA, MS, MBA, PhD, JD, MD etc.)

All of these questions allow participants to choose "prefer not to answer" as an option. However, participants who select it will not qualify for the study and any identifying data is deleted.

After participants provide the needed demographics information, GSA checks the current sample distribution at the demographic level and determines whether the quota for the participant's self-asserted demographic group has been filled. If the quota has already been met the participant will not qualify for the study and any identifying data is deleted.

4.3 Communications with subjects

All communications are done electronically:

Website: GSA built a website for the study that explains why the study is being done, has details of all the data that is being collected and why, as well as an expansive list of "Frequently Asked Questions." See the PDF document titled "Landing Page Content - Identity Proofing Equity Study" for details.

Social Media: GSA will use Facebook and Instagram ads that are targeted to the needed demographics. Mockups for these ads can be found in the document titled “GSA Identity Proofing Equity Study Marketing.”

Email Newsletters: GSA is partnering with multiple organizations that work directly with the communities and populations that GSA needs to recruit for the study. The plan is to post advertisements in their regularly scheduled email newsletters. These organizations include:

- LatinX in AI (LXAI)
- Black Tech Pipeline
- Tribaja
- Inclusively

GSA will continue to contact similar organizations that work with the needed populations as the recruitment process progresses.

Direct Email:

- Link to Study Screening Survey
- Completion Confirmation (sent by GSA)
- Compensation instructions (sent by Rekrewt through Tremendous.com)
- Troubleshooting conversations

5. Inclusion and Exclusion Criteria

Participants must

- be over 18 years of age,
- agree to the terms and conditions of the study,and
- answer all the required demographic questions

Furthermore, as the recruitment progresses, potential participants may be excluded from the study if the study has already reached the necessary number of participants from their specific demographic group.

6. Number of subjects and demographic information

The study aims to recruit 4,000 participants across race/ethnicity, gender, age, income, and education:

1. A minimum of 770 testers/group for each of the following racial/ethnic groups:
    - White/Caucasian
    - African American
    - Hispanic/Latinx
    - Asian, Hawaian, or/ Pacific Islander
    - American Indian or Alaskan Native

2. A minimum of 1285 testers for each of the following gender identities:
    - Male
    - Female
    - Another gender identity
3. A minimum of 300 testers for each of the following Age Ranges
    - 18 to 25
    - 26 to 40
    - 41 to 55
    - 56+
4. A minimum of 300 testers for each of the following yearly Household Income Range
    - $0- $24,999 / year
    - $25,000 - $75,999 / year
    - $76,000 - $99,999 / year
    - $100,000 and higher / year
5. The recruitment service shall recruit a minimum of 300 testers for each of the following Education Levels:
    - Some high school or less
    - High school diploma or GED graduate
    - Some college or university, but no degree
    - Bachelor's degree
    - Graduate or professional degree (MA, MS, MBA, PhD, JD, MD etc.)

7. Where will the research be conducted

Data is collected over the internet and requires that participants have a mobile device with internet access.

The U.S. General Services Administration, headquartered at 1800 F St NW, in Washington DC is conducting the study. The Center for Identity Technology Research (CITeR) and Clarkson University, located at 8 Clarkson Ave. in Potsdam, NY, will perform the statistical data analysis.

8. Study timeline

The study timeline is dependent upon meeting the recruitment target of 4,000 participants across various demographics. GSA plans to begin participant recruitment no later than one month after IRB approval. GSA estimates a three-month recruitment timeline. The data analysis and writing of the study results will take an additional 6-8 months. Adding an additional 6 month estimate for peer review yields a cumulative study timeline of up to two years after IRB approval.

9. Data

9.1 Where will data be stored?

GSA will retain records of this study in accordance with GSA's records schedule for Customer Research and Reporting Records and any other applicable federal records schedules. GSA's Customer Research and Reporting Records schedule requires certain research records to be destroyed six years after the end of the fiscal year when collection is complete.

The data collected will be stored in an encrypted GSA Google Drive that can only be accessed by a tightly controlled group of GSA civilian and contractor personnel.

A de-identified dataset (with all PII removed) will be encrypted and shared with CITeR for data analysis.

9.1 How will the data be analyzed?

GSA is partnering with CITeR to conduct the analysis of the de-identified data. CITeR is developing statistical methods to determine if there is a meaningful difference between groups or if that difference is only due to random error and chance.

This study will incorporate statistical methods to measure fairness by examining false negatives. Biometric solutions used widely by the public are typically based on "verification" or one-to-one matching. A false negative error is when the correct individual is falsely rejected: e.g. the system does not match a participant's selfie with the provided government identification. This "error" may block an individual from accessing benefits which they are entitled to. The number of demographic groups being compared impacts the variation as an increased number of groups increases the chances that a difference between groups may be found "by chance," and thus adjustments need to be made in the test due to this effect, often called multiplicity[14].

To fulfill the purpose of this research, GSA requires sufficient representation from each of the demographic groups of interest. GSA will use quota sampling to achieve this; quota sampling is similar to stratified sampling in that sampled individuals will be placed into their respective groups. Rather than sampling from subgroups as in stratified sampling, quota sampling continues to sample randomly until the specified number of individuals in each group is met.

GSA will apply quota sampling to split up the population based on demographic attributes. Specifically, the team will be examining various identity proofing results to study the correlation between false negatives and demographics.

The team will analyze a minimum of 300 participants per group. From a statistical point of view, this provides more subjects within a factor (e.g., age only) to ensure differences between groups (e.g., differences between 18-25 and 56+) can be determined.

[14] J. Hsu, *Multiple Comparisons: Theory and Methods*. Chapman & Hall/CRC, 1996.

Each respondent will submit their information into multiple identity proofing products so that the team can assess the following identity-proofing false negatives:

- Device risk failure rate - Proportion of devices that fail device risk check
- Document false reject rate - Proportion of genuine document authentication transactions with truthful claims of an authentic document that are incorrectly denied
- Document failure to acquire - Proportion of document authentication attempts for which the system fails to capture or locate an image or signal of sufficient quality
- False non-match rate (ISO 2382-37) - Proportion of the completed biometric mated comparison trials that result in a false non-match
- False reject rate (ISO 19795-1) - Proportion of biometric verification transactions with truthful claims of identity that are incorrectly denied
- Bonafide Presentation Classification Error Rate (ISO 30107-3) - Proportion of bona-fide biometric presentations incorrectly classified as presentation attacks in a specific scenario
- Personally Identifiable Information (PII) failure rate - Proportion of PII check which fail to match an authoritative source
- Phone verification failure rate - Proportion of OTP requests which fail to be completed
- Overall false negative rate - document authentication + selfie verification + liveness + PII validation + device possession check

GSA will follow International Organization for Standardization (ISO) standards for biometric performance and perform statistical analysis after the test is run. Statistics:

- Analysis based on Rule of 3 (ISO 19795-1)
- Simulations of bootstrapping (ISO 19795-1)

For example, for a test set of 385 subjects, following the “Rule of 3”, if there are zero errors, the upper bound of the 95% confidence interval for the False Rejection Rate (FRR) is 0.7%. This should be sufficient since, as it is expected that all the identity-proofing products being used in the instrumentation collection will have an FRR of greater than 1%. However, this analysis needs to consider whether cross factor analysis is also needed, e.g., age and educational level. The analysis will be repeated using bootstrapping where zero errors are not assumed to determine the confidence interval that would be obtained for various scenarios.

10. Subject Privacy

How will privacy be protected?

GSA will be collecting the following information from participants:

1. **Name and email** - Once a user agrees to the terms and conditions of the study to determine eligibility. This information will also be provided to Rekrewt to initiate compensation for users that complete the study.
2. **Demographic data** - Demographic information will be provided to Rekrewt to tailor social media ads to garner interest from lower demographic groups (See question #6).
3. **Biometric and PII data** - The selfie, image capture of the front and back of driver's license, SSN, and phone number will be passed to the identity proofing vendors for evaluation and then stored on the GSA Google Drive.
4. **Self-asserted skin tone** - Participants choose the value from the Monk Skin Tone scale that best matches their complexion. *Answering this question is optional.*
5. **Usability evaluations** - Participants are asked to rate and provide feedback at various points in the study. *Answering these questions is optional.*
6. **Consent to share redacted images** - Participants are asked for permission in using their redacted images in the published results. For example, these images can be used to explain failures where the identity card is damaged or the pictures taken by participants are blurry. Participants can decline to share this without jeopardizing their compensation.

GSA acknowledges that there are a multitude of privacy concerns with this collection. Therefore the design of the study platform ensures that any PII collected will be safeguarded in a separate dataset and secured according to GSA's Security and Privacy policies. GSA will also publish a publicly available Privacy Impact Assessment (PIA) that explains how both the study platform and all commercial components and entities are safeguarding participants' privacy.

GSA requires all information systems that are publicly accessible to attain an "Authority to Operate" (ATO). The security measures described below are required to fulfill GSA's ATO requirements:

- All data collected will be maintained within authorized GSA information systems as detailed by the relevant Systems of Records Notice, GSA/TTS-1. The authorization process for these systems includes the testing of security controls, scanning of the system for vulnerabilities, and manual penetration testing.
- All data collected as part of the equity study is encrypted at all times.
- Only GSA project staff that have passed background checks, and have a need to know will be provisioned access to the data.
- All user accounts undergo an annual review to ensure compliance with account management requirements.

All vendors are undergoing an IT security requirements review to ensure that their systems and policies comply with GSA guidelines.

To ensure transparency and obtain informed consent, GSA has added steps in the study platform where participants have the opportunity to read the relevant privacy policies (from both GSA and the identity-proofing vendors). GSA has also provided a Privacy Act Statement and privacy policies of the Identity Proofing (IdP) vendors that are part of the study for all participants to review. Participants must assert that they understand and accept these policies before they are allowed to access the study platform. This step will help participants understand how their information is being used, stored, and protected.

All PII will be stored within the GSA Google Drive and is only accessible by a limited number of project team members. The data analysis will only be done using the de-identified dataset provided to CITeR. None of the Identity Proofing vendors are allowed to retain any user data

The final peer-reviewed publication will only contain de-identified or redacted data. Results will be presented in graphs and in groupings. The publication will only contain demographic information and de-identified study results. The following processes will be leveraged to de-identify data:

- participant information is masked,
- participant names, date of birth, social security numbers, addresses, and phone numbers are suppressed,
- no direct identifiers will be included

11. Financial Compensation

Is there any financial compensation to subjects? If so, what is provided and how is it given?

All participants who successfully complete the study will be compensated with a $25 e-gift card that the recruitment agency (Rekrewt) will be issuing via Tremendous.com.

GSA will provide Rekrewt a daily list of the name and email addresses of participants who have completed the study.

12. Consent

How do subjects give consent (written, verbal, digital?)

All participants will consent digitally when they register online for the study. Consent is split into three components. The first half of the consent collection happens during the participant screening process. All *identifying data* from participants who do not consent or are not eligible for the study is deleted. However, GSA will retain non-eligible participants' demographic information (devoid of name, and other identifying data) to analyze trends in study abandonment

and recruitment challenges as noted in the Rules of Use: “If you [the participant] decide to terminate your participation after completing the demographics survey, GSA will retain the de-identified demographic information to study drop-off rates and the recruitment process.“ The Rules of Use can be found in Attachment 1 in the document titled “GSA Identity Equity Study - Consent Forms and Attachments".

The second part of the consent collection only applies to eligible participants and is collected “just in time” before the biometric information collection begins. The third consent statement is collected by Socure, one of GSA’s identity proofing vendors; this third statement is redundant but was required by Socure’s legal team.

Consent agreements are stored within the GSA Google Drive system as part of the study’s overall dataset. This data also allows GSA to assist users in the event that they submit a request regarding the use of their data. Furthermore, in the event that a participant decides to revoke consent after completing the study, GSA will quarantine their data to ensure it is not used in future research.

See Appendix 1 for the consent documents provided to participants.

12.1 Broad Consent Request

The Privacy Act Statement notes that “The PII you [the participant] provide in this Equity Study will only be used for the purpose of this study and future Equity Studies if applicable.” This is intended to say that GSA can use the data to examine multiple equity concerns and questions.

However, the identifiable data will not be shared with other third parties or government agencies. The Privacy Act Statement can be found as Attachment 2 in the document titled “GSA Identity Equity Study - Consent Forms and Attachments".

Please let us know if this means we need to request broad consent and if so what changes are needed to our consent documentation.

Appendix 1 - Consent documents

Consent #1 - Pre screening (2 pages)

GSA

Study Purpose and Participation Requirements

GSA will publish a de-identified, peer-reviewed report assessing the impact of ethnicity, race, gender, income, and other demographic factors on identity-proofing tools. You will test and provide feedback on multiple identity-proofing components and products including:

1. Document Authentication

- You will take and submit pictures of your identity document (e.g., a driver's license, state ID)
- You will take and submit a "selfie" (a picture of your face)
- You will rate the "ease of use" of the tool
- You will repeat these steps for **FIVE (5)** different document authentication products

2. Personal Information Validation

You will provide your:

- Full legal name
- Date of birth
- Physical address
- Social security number
- Phone number

3. Device Checks

- You will receive and confirm a security code by phone (voice call or text message)
- This application will scan your mobile device and evaluate it for its capabilities, features, and uses.

4. Exit Survey

- You will answer three (3) short questions

Participant Acknowledgements and Agreements

With your consent, GSA will share your information with third party vendors who will only use it to verify your information. GSA will collect and store your data along with the vendors' validation and verification results. None of your data will be used for marketing or purposes other than this research.

GSA will instruct the vendors to delete your data from their systems within 24 hours of collection.

GSA will share an aggregated de-identified dataset with CITeR and Clarkson University for analysis. Your personally identifiable information (PII) will not be included in the data that GSA shares with GSA's research partners.

Your participation is voluntary. You can withdraw from the study or cancel your permission for GSA to use your identifying information prior to study completion by closing the study webpage on your browser or contacting the GSA researchers at identityequitystudy@reseach.gsa.gov or (202) 969-0772.

If you decide you don't want your information used, or if you have any questions or complaints, you may also contact a person not on the research team at the Biomedical Research Alliance of New York Institutional Review Board at (516) 318-6877 or at www.branyirb.com/concerns-about-research. Information that was already collected may still be used.

Risk and Harm: GSA and its partners are committed to protecting your data to the greatest extent. However, there is always the risk of loss of confidentiality of your personal information used for this study. If this were to happen, GSA will promptly inform you with additional information.

* Please read the study's Rules of Use and the Privacy Act Statement.

This requires a response

☐ I have reviewed the Rules of Use and Privacy Act Statement and agree to abide by them.

* Are you interested in participating in the study?

This requires a response

○ I am interested in participating in the study. I consent to the collection and use of my information as described above.

○ I am not interested in participating in the study. I do not consent.

Previous Next

* Please provide your contact information. (This information will be used to notify the recruitment partner when you have completed the study, so they can follow up with you about compensation.)

This requires a response

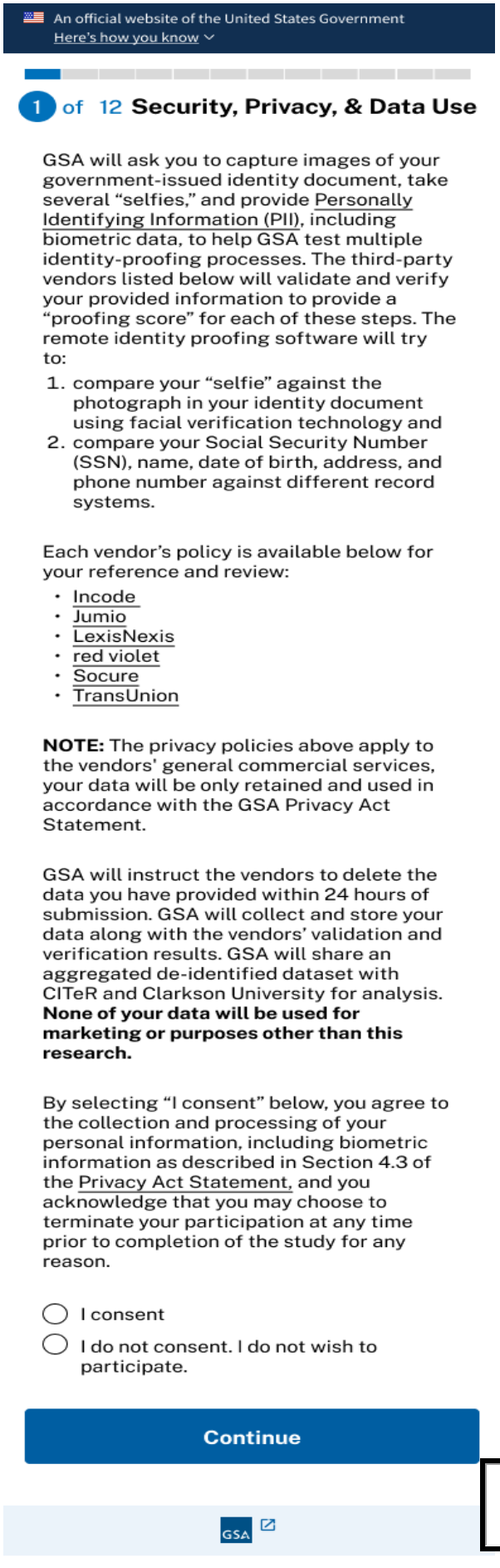
First Name

Last Name

Email Address

Submit

Consent #2 - Post-screening



Consent #3 - Vendor Specific