How to Use Perplexity for Data on Women in Combat Roles

Objectives

Our **immediate goal** is to collect cross-national data on women in military institutions. The **down-stream analytical goal** is to better understand how women's participation in combat, deployment to combat zones, and service in military institutions affects the conduct of war and conflict outcomes.

Our work will entail extending a partially-completed dataset on Women in Conflict Roles (developed in 2015-2017 by Jessica Trisko-Darden and Laia Balcells). We will focus on collecting data on the following **sample of countries** in wartime, during the period 1939-2023:

1. Argentina (1982)
2. Armenia (1993)
3. Australia (1939, 1940, 1941, 1942, 1943, 1944, 1945, 1950, 1951, 1965, 2001, 2003)
4. Azerbaijan (1993)
5. Belgium (1940, 1944, 1945, 1951, 1999)
6. Bosnia & Herzegovina (1992)
7. Bulgaria (1941, 1944, 1945)
8. Cambodia (1970, 1977, 1978, 1979)
9. Canada (1939, 1940, 1941, 1942, 1943, 1944, 1945, 1950, 1951, 1953, 1991, 1999, 2001, 2003)
10. Chad (1986)
11. China (1941, 1942, 1944, 1950, 1951, 1953, 1954, 1958, 1962, 1979, 1987, 2020)
12. Colombia (1951, 1953)
13. Congo - Kinshasa (1975)
14. Croatia (1992)
15. Cuba (1975, 1977)
16. Cyprus (1974)
17. Denmark (1999)
18. Egypt (1948, 1956, 1967, 1969, 1973, 1991)
19. El Salvador (1969)
20. Eritrea (1998, 1999, 2016)
21. Ethiopia (1940, 1941, 1951, 1953, 1977, 1998, 1999, 2016)
22. Finland (1939, 1941)
23. France (1940, 1942, 1944, 1945, 1950, 1951, 1956, 1958)
24. Gabon (1940)
25. Germany (1939, 1940, 1941, 1942, 1943, 1944, 1945, 1999, 2001)
26. Greece (1940, 1941, 1943, 1944, 1945)
27. Honduras (1969)
28. Hungary (1941, 1942, 1944, 1956, 1999)
29. India (1947, 1948, 1962, 1965, 1971, 1999, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020)
30. Iran (1980, 1984, 1985, 1986, 2018, 2019, 2020, 2021, 2022, 2023)
31. Iraq (1941, 1948, 1973, 1980, 1984, 1985, 1986, 1987, 1990, 1991, 2003)
32. Israel (1948, 1956, 1967, 1969, 1973, 1982, 2018, 2019, 2020, 2021, 2022, 2023)
33. Italy (1940, 1941, 1942, 1943, 1991, 1999)
34. Japan (1939, 1941, 1942, 1943, 1944, 1945)
35. Jordan (1948, 1967, 1973)
36. Kuwait (1990)
37. Kyrgyzstan (2021, 2022)
38. Laos (1968, 1971)

39. Lebanon (1948)
40. Libya (1979, 1986)
41. Mongolia (1939, 1945)
42. Morocco (1957, 1991)
43. Netherlands (1940, 1942)
44. New Zealand (1939, 1940, 1942, 1943, 1944, 1945, 1950, 1951)
45. North Korea (1950, 1951, 1953)
46. Norway (1940, 1999)
47. Oman (1991)
48. Pakistan (1947, 1948, 1965, 1971, 1999, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020)
49. Peru (1995)
50. Poland (1939, 1940, 1942, 1943, 1944, 1945, 2003)
51. Qatar (1991)
52. Romania (1941, 1942, 1944)
53. Russia (1939, 1941, 1942, 1943, 1944, 1945, 1956, 2022, 2023)
54. Saudi Arabia (1973, 1991)
55. Senegal (1940)
56. Somalia (1977)
57. South Africa (1939, 1940, 1942, 1943, 1944, 1945, 1975)
58. South Korea (1950, 1951, 1953, 1965, 1966)
59. South Sudan (2012)
60. Spain (1957, 1999)
61. Sudan (2012)
62. Syria (1948, 1967, 1973, 1982, 1991)
63. Taiwan (1954, 1958)
64. Tajikistan (2021, 2022)
65. Tanzania (1978, 1979)
66. Thailand (1940, 1950, 1951, 1953, 1967, 1970)
67. Turkey (1950, 1974, 1999)
68. Uganda (1978)
69. Ukraine (2022, 2023)
70. United Arab Emirates (1991)
71. United States (1941, 1942, 1943, 1944, 1945, 1950, 1951, 1953, 1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1991, 1999, 2001, 2003, 2020, 2021, 2022, 2023)
72. Vietnam (1965, 1966, 1967, 1968, 1969, 1970)
73. Yugoslavia (1941, 1992, 1999)

For each country and each year on this list, we will use the following **coding protocol**:

- Q1. Were women allowed to serve in the military in [COUNTRY] in [YEAR]? This is typically identified in the country's constitution or other relevant body of law.
- Q2. Was it mandatory for women to serve in the military in [COUNTRY] in [YEAR]?
- Q3. Were women subject to conscription in [COUNTRY] in [YEAR]?
- Q4. Were there female-only units in [COUNTRY] in [YEAR]?
- Q5. Were there mixed-sex military units in [COUNTRY] in [YEAR]? These are units where women serve side by side with men.
- Q6. Were there women in military leadership positions in [COUNTRY] in [YEAR]? If yes, list the highest rank.
- Q7. Did women serve in official combat roles in [COUNTRY] in [YEAR]? This includes positions whose duties entail direct engagement with enemy forces (e.g. infantry, combat engineers, armor crew, snipers, fighter pilots, forward-deployed drone operators). If yes, describe.
- Q8. Were female service members deployed to combat zones by [COUNTRY] in [YEAR]? This includes serving in proximity to ongoing military operations, but not necessarily direct participation in combat (e.g. logistics, medicine, administration). If yes, describe.
- Q9. Were there female military personnel in [COUNTRY] in [YEAR]? If yes, record the annual number.

- Q10. Were there women in the Army/Ground Forces in [COUNTRY] in [YEAR]? If yes, record the annual number.
- Q11. Were there women in the Air Force in [COUNTRY] in [YEAR]? If yes, record the annual number.
- Q12. Were there women in the Navy in [COUNTRY] in [YEAR]? If yes, record the annual number.
- Q13. Were there women in the Marines/Naval Infantry in [COUNTRY] in [YEAR]? If yes, record the annual number.
- Q14. Were there women in police under military jurisdiction (i.e. where a branch of the police is formally integrated into the military command structure) in [COUNTRY] in [YEAR]? If yes, record the annual number.
- Q15. Were there women in other military branches in [COUNTRY] in [YEAR]? If yes, record the annual number.
- Q16. Were there women working for the military as civilian, contractor or auxiliary personnel (not enlisted or commissioned) in [COUNTRY] in [YEAR]?
- Q17. Were female civilian, contractor or auxiliary personnel (not enlisted or commissioned) deployed to combat zones by [COUNTRY] in [YEAR]?

Your job will be to feed this coding protocol as a series of queries to an AI chatbot, and to vigilantly check and document the sources used by the model to substantiate its responses. In particular, you will use Perplexity.ai for this purpose.

Perplexity.ai is among the AI research tools that Georgetown recommends, and it stands apart for its transparency in providing URLs and other source information for its results. Perplexity.ai also lets users choose from multiple industry-standard models (including GPT4, Gemini, and DeepSeek on U.S.-based servers), along with its in-house models (e.g. Sonar). These features will allow us to validate each set of results against those from other AI models and sources, and to partially account for classification uncertainty. As students, you are also eligible for a free month of the "Pro" tier, which we need for this project.

You should think of this exercise as being akin to conducting a series of "expert interviews". These experts are highly knowledgeable, but they sometimes disagree, because – like humans – they read different sources and think differently. They also sometimes make mistakes, and it is not always clear which answer is correct and which is wrong. This is why we will

- a) *Check their work*, by requesting full bibliographic information for each set of answers.

- b) *Get a second opinion*, by asking a different AI "expert" an identical set of questions.

Specifically, we will be using six "flavors" of Perplexity.ai's search, for each country-year:

1. **Perplexity "Auto" search, with web and academic sources**. Perplexity's "Auto" search utilizes an ensemble of models, including its proprietary LLMs (`pplx-7b-online`, `pplx-70b-online`, which are fine-tuned versions of open-source models like `Mistral-7b` and `Llama2-70b`), and external models like OpenAI's GPT-4 and Anthropic's Claude 3. It prioritizes speed over deep reasoning by using lightweight computations and summarizing information from (mainly) indexed sources.
2. **Perplexity "Auto" search, academic sources only**. Same as above, but using only materials derived from scholarly databases, peer-reviewed journals, scientific publications, and other reputable academic resources.
3. **OpenAI's `o3-mini`, with web and academic sources**. This model is designed for advanced reasoning tasks, like solving STEM problems, coding challenges, and scientific inquiries, and integrates real-time web search to provide up-to-date answers.
4. **OpenAI's `o3-mini`, academic sources only**.
5. **DeepSeek's `R1`, with web and academic sources**. DeepSeek's model for analytical and research-intensive tasks, which has a different architecture from `o3-mini`. Perplexity hosts this model on U.S. servers, which helps ensure legal data and privacy protections, and mitigates risks of unauthorized access or surveillance by foreign entities like China.
6. **DeepSeek's `R1`, academic sources only**.

For each country and year, and each model/search, the **deliverables** include 2 files:

1. Response to coding protocol as a Markdown file
    - file name: `[ISO3 country code]_[year]_[model]_[types of sources].md`
2. Bibliography as a BibTeX file
    - file name: `[ISO3 country code]_[year]_[model]_[types of sources].bib`

ISO3 country codes are listed in the spreadsheet `wicr_tracker.csv` in `YZRA/Data/WICR/` (or on the web: `en.wikipedia.org/wiki/List_of_ISO_3166_country_codes`). For model names, we will simply put "Perplexity" (for "Auto" search), "OpenAI" (for `o3-mini`) or "DeepSeek" (for `R1`). For "types of sources" we will use "webaca" to denote "web and academic sources" and "aca" for "academic only".

There should be 12 files for each country-year. For example, for Argentina in 1982:

- `ARG_1982_Perplexity_webaca.md`
- `ARG_1982_Perplexity_webaca.bib`
- `ARG_1982_Perplexity_aca.md`
- `ARG_1982_Perplexity_aca.bib`
- `ARG_1982_OpenAI_webaca.md`
- `ARG_1982_OpenAI_webaca.bib`
- `ARG_1982_OpenAI_aca.md`
- `ARG_1982_OpenAI_aca.bib`
- `ARG_1982_DeepSeek_webaca.md`
- `ARG_1982_DeepSeek_webaca.bib`
- `ARG_1982_DeepSeek_aca.md`
- `ARG_1982_DeepSeek_aca.bib`

These files are provided as examples in the `YZRA/Data/WICR/Processed` folder.

The following instructions explain the workflow step-by-step.

---

1. Setting Up a Free Account with Perplexity.ai

1. **Visit the Website**: Navigate to Perplexity.ai.
2. **Click "Sign Up"**: Look for the blue "Sign Up" button on the homepage.
3. **Choose Sign-Up Method**:
    - Use your university email (e.g., `netid@georgetown.edu`) by selecting "Continue with Email."
    - NOTE: While you can also sign up via Google or Apple accounts, students with verified `.edu` addresses are automatically upgraded to the "Perplexity Pro" tier for one month or longer when they sign up, and are eligible for a discounted rate of $4.99/month when the trial period expires.
4. **Verify Email**: Check your inbox for a verification email and follow the link to confirm your account.
5. **Log In**: Once verified, log in to Perplexity.ai.

---

2. Submit First Query: Coding Protocol

1. **Pick a country and year**
    - Use our Trello board to claim a set of country-years, and drag the card to the "Working" column
    - This list is also in the spreadsheet `wicr_tracker.csv`
2. **Open Perplexity**
    - Log in to your Perplexity account.
    - On the left-hand menu, click **Start New Thread**.
3. **Enter Your First Query**:
    - You will enter two queries back-to-back. The first will ask the model to respond to the coding protocol. The second will ask the model to provide a bibliography. Text for these prompts is in the plain text file `coding_protocol.txt` in the directory `YZRA/Data/WICR/Raw`.

- For the first of these queries ("coding protocol"), you will submit the list of 17 questions from above, with a short preamble and instructions to explain the task to the chatbot.
- Copy the following text from `coding_protocol.txt` into the search box, filling in the blanks in the first sentence (e.g. "replacing [COUNTRY] with Argentina and [YEAR] with 1982").

```
Please answer the following questions, replacing [COUNTRY] with ___ and [YEAR]
with ___. Keep your answers brief (Y/N), unless prompted to describe. In the latter
case, keep your response to one sentence. If you have insufficient information to
answer a particular question, leave its answer blank.

Q1. Were women allowed to serve in the military in [COUNTRY] in [YEAR]? This is
    typically identified in the country's constitution or other relevant body of law.
Q2. Was it mandatory for women to serve in the military in [COUNTRY] in [YEAR]?
Q3. Were women subject to conscription in [COUNTRY] in [YEAR]?
Q4. Were there female-only units in [COUNTRY] in [YEAR]?
Q5. Were there mixed-sex military units in [COUNTRY] in [YEAR]? These are units where
    women serve side by side with men.
Q6. Were there women in military leadership positions in [COUNTRY] in [YEAR]? If yes,
    list the highest rank.
Q7. Did women serve in official combat roles in [COUNTRY] in [YEAR]? This includes
    positions whose duties entail direct engagement with enemy forces (e.g. infantry,
    combat engineers, armor crew, snipers, fighter pilots, forward-deployed drone
    operators). If yes, describe.
Q8. Were female service members deployed to combat zones by [COUNTRY] in [YEAR]?
    This includes serving in proximity to ongoing military operations, but not
    necessarily direct participation in combat (e.g. logistics, medicine,
    administration). If yes, describe.
Q9. Were there female military personnel in [COUNTRY] in [YEAR]? If yes, record the
    annual number.
Q10. Were there women in the Army/Ground Forces in [COUNTRY] in [YEAR]? If yes,
     record the annual number.
Q11. Were there women in the Air Force in [COUNTRY] in [YEAR]? If yes, record the
     annual number.
Q12. Were there women in the Navy in [COUNTRY] in [YEAR]? If yes, record the
     annual number.
Q13. Were there women in the Marines/Naval Infantry in [COUNTRY] in [YEAR]? If yes,
     record the annual number.
Q14. Were there women in police under military jurisdiction (i.e. where a branch of
     the police is formally integrated into the military command structure) in
     [COUNTRY] in [YEAR]? If yes, record the annual number.
Q15. Were there women in other military branches in [COUNTRY] in [YEAR]? If yes,
     record the annual number.
Q16. Were there women working for the military as civilian, contractor or auxiliary
     personnel (not enlisted or commissioned) in [COUNTRY] in [YEAR]?
Q17. Were female civilian, contractor or auxiliary personnel (not enlisted or
     commissioned) deployed to combat zones by [COUNTRY] in [YEAR]?

Please provide your responses in a list like this:

- Q1. Y
- Q2. N
```

```
etc.
```

   - It is very important that you **do not modify the text** of this query, apart from the country and year. Otherwise, the results are not truly comparable across searches. If something is unclear, you should clear it up in follow-up questions, rather than by editing this query.
   4. **Set Model and Sources**:
      - Ensure **Auto Mode** is selected (default setting).
      - Click on the "Filters" icon below the search bar.
      - Select **Web and Academic Sources Only**.
   5. **Submit Query**: Press "Enter" or click the search icon to start the thread.

---

3. Export Answer as a Markdown File

   1. **Read the Response**:
      - After you submit the query, a new page will open, with your query followed by a the model's response.
      - This response should be a list like "Q1: N, Q2: N, etc.", sometimes followed by a few notes.

   2. **Locate Export Option**:
      - Scroll to the bottom of the response box.
      - Find and click on the "Export Answer" button (usually labeled with an export icon).
   3. **Save Markdown File**:
      - Choose Markdown format.
      - Save the text file to the directory `YZRA/WICR/Data/Processed`
      - Name the file like this: `[ISO3 country code]_[year]_Perplexity_webaca.md`

---

4. Submit Second Query: Bibliography

   1. **Locate the "Follow Up" Question Box**
      - Find the search box below the current response. This is a box to ask follow-up questions to refine or expand on existing answers (unlike new threads, which start a fresh conversation unrelated to previous responses).
   2. **Enter Your Second Query**:
      - Now you will ask the model to provide a bibliography for its previous response. Text for this prompt is in the plain text file `coding_protocol.txt` in the
      - Copy the following text from `coding_protocol.txt` into the search box:

```
Please generate a .bib file with bibliographic entries (including DOIs and URLs, where
available) for academic, government or journalistic sources that can substantiate each answer.
```

   - After you submit the query, a new response box will appear, below the first one.
   - This response should be a formatted document with bibliographic entries (e.g. usually beginning with tags like `@online{}` or `@article{}`), sometimes followed by a few notes.

   3. **Copy the Text**:

- Copy the text from *inside the text box only* (omitting any additional comments) through either the "Copy" button (clipboard icon, or with keyboard shortcuts as described in 3.2.
4. **Paste into a Text File**:
    - Open a plain text editor
    - Paste using `Ctrl+V` (Windows/Linux) or `Cmd+V` (MacOS).
5. **Save the .bib File**
    - Save the text file to the directory `YZRA/WICR/Data/Processed`
    - Name the file like this: `[ISO3 country code]_[year]_Perplexity_webaca.bib`

---

5. Ask Follow-Up Questions (optional)

1. **Check the Responses for Obvious Errors and Inconsistencies**
    - These models sometimes commit errors, especially when responding to complex, nuanced queries. Answers may also be logically inconsistent (e.g., "N" to Q1 but "Y" to Q2).
    - Bibliographies may also be incomplete, or improperly formatted.
    - You can clear some of this up by asking more follow-up questions.
2. **Clarify or Expand**:
    - Ask for additional citations, explanations, or clarification of inconsistencies in the same thread.
    - For example: "Doesn't your answer to Q2 contradict your answer to Q3?"
3. **Reference Previous Responses**:
    - Use phrases like "Based on your previous answer to Q1" to maintain continuity within the thread.
    - For example: "In your earlier answer, you mentioned X. Could you elaborate further?"
    - Ensure you are typing in the same thread window by *not* clicking "New Thread."
4. **Save the Follow-Ups**
    - Export follow-up query-answer pairs as Markdown files, as described in 4.
    - Name the new files like this (in `YZRA/Data/WICR/Processed`):
        - `[ISO3 country code]_[year]_Perplexity_webaca_followup1.md`,
        - `[ISO3 country code]_[year]_Perplexity_webaca_followup2.md`, etc.
    - Be sure not to overwrite any of the other files you created for the country-year.

---

6. Repeat Steps 2-5 with Academic Sources Only

1. **Start a New Thread**, as described in Step 2, but *keeping the same country and year*.
2. In Step 2.4, select only **Academic Sources** in the source filters menu.
3. Repeat Steps 2-5 as described, but modifying the file names as follows to reflect the change:
    - `[ISO3 country code]_[year]_Perplexity_aca.md`
    - `[ISO3 country code]_[year]_Perplexity_aca.bib`

---

7. Repeat Steps 2-6 with OpenAI's o3-Mini Model

1. **Start a New Thread**, as described in Step 2, but *keeping the same country and year*.
2. **Switch Model**:
    - In Step 2.4, change model to "Reasoning with o3-Mini" using the model selector at the top of Perplexity's interface.
    - Note that this model runs much more slowly than the "Auto" search.
3. Repeat Steps 2-6 using this model, running this once with web+academic sources and once with academic sources only.
4. Four new files should be generated at the end of this process:
    - `[ISO3 country code]_[year]_OpenAI_webaca.md`

- `[ISO3 country code]_[year]_OpenAI_webaca.bib`
- `[ISO3 country code]_[year]_OpenAI_aca.md`
- `[ISO3 country code]_[year]_OpenAI_aca.bib`

---

8. Repeat Steps 2-6 with DeepSeek's R1 Model

1. **Start a New Thread**, as described in Step 2, but *keeping the same country and year*.
2. **Switch Model**:
   - In Step 2.4, change model to "Reasoning with R1" using the model selector at the top of Perplexity's interface.
   - Note that this model runs even slower than `o3-mini`.
3. Repeat Steps 2-6 using this model, running this once with web+academic sources and once with academic sources only.
4. Four new files should be generated at the end of this process:
   - `[ISO3 country code]_[year]_DeepSeek_webaca.md`
   - `[ISO3 country code]_[year]_DeepSeek_webaca.bib`
   - `[ISO3 country code]_[year]_DeepSeek_aca.md`
   - `[ISO3 country code]_[year]_DeepSeek_aca.bib`

Additional Tips

Tip 1: Keep Everything Updated

1. Keep track of your and the team's progress with the spreadsheet `wicr_tracker.csv` in the `YZRA/Data/WIC` folder. When you begin working on a file (and have created a working copy), change the value in the `working` column for that file from `N` to `Y`. Save and close. Then do the same for the `processing` column when you're done.
2. Also keep things up to date on our Trello board. Move the card for each file from `To Do` to `Doing` and `Done` as you go.

Tip 2: Save your threads

- Don't delete your past threads in Perplexity's Library, in case you ever need to go back and check or modify something.