Data Plan:

Data Elements:

- Ultrasound images: 2D ovarian ultrasound scans (grayscale, annotated).
- Patient demographics: Age (numerical feature).
- Biomarkers: CA-125 and BRCA (binary).
- These elements are combined to build multimodal inputs (image + tabular) for the prototype.

Bias Considerations:

- **Class collapse:** Original dataset has 8 sub-classes of tumors.
  For this prototype, it was collapsed into two groups, benign (0) and malignant (1). While this simplifies the task, it may hide important differences across tumor subtypes.

- **Sampling bias:** The dataset is relatively small, so splits (train/val/test) may not fully represent real-world diversity. Balanced sampling was used to reduce skew, but further validation on larger, external datasets is needed.

- **Overfitting risk:** Small dataset size increases the chance of overfitting. It can be mitigated with careful training/validation splits and regularization.

Compliance & Intended Use:

- This prototype is for research and educational purposes only.

- It is not clinically validated and must not be used for patient care or real-world diagnostic decisions.

- Any future clinical application would require regulatory approval, multi-institutional data collection, and adherence to data privacy standards.