Stage 2: EXPLAIN deliverable summary

Deadline: March 16, 2026 (?)

Goal

Stage 2 asks teams to justify the scientific basis of their materials problem and, crucially, demonstrate that the project is feasible given the data they can realistically obtain this semester.

What students must deliver (emphasis on data)

- Data acquisition and "data landscape" assessment (major focus)
  - Identify data sources: databases (e.g., Materials Project/AFLOW), published datasets, repositories, or literature extraction.
  - Specify what will be collected: target property(ies), input features (composition, processing, structure descriptors), units, metadata, and any required experimental/simulation conditions.
  - Confirm accessibility: API availability, download format, licensing/usage constraints, and whether the team can access everything needed.
  - Estimate quantity and coverage: expected number of samples, design-space coverage, and whether the data distribution matches the project's goals.
  - Assess quality: missing values, duplicates, inconsistent units/conditions, uncertainty/noise, measurement or computation fidelity, and potential biases.
  - Define a concrete acquisition + cleaning plan: how data will be pulled, merged, standardized, filtered, and documented for reproducibility.

    - Identify gaps early: what is missing, how critical it is, and whether the team can reasonably fill gaps (e.g., additional computations, simplified surrogate targets, narrower scope).
- Feasibility statement anchored to data
    - A clear “yes/no/yes-with-scope-reduction” conclusion on feasibility, justified by the data landscape.
    - If data is insufficient: define a scope change or an alternative plan that still supports a meaningful analysis/design outcome.
- Data-driven risk analysis
    - List key data risks (availability, insufficient volume, poor quality, missing key features, inconsistent conditions, bias).
    - Estimate likelihood and impact, and outline mitigation (backup datasets, feature proxies, narrowing the design space, alternative targets, sanity checks).

Suggested report structure (simple)

1. Brief recap of Stage 1 problem framing
2. Materials science foundation (why the problem and variables make sense)
3. Data acquisition plan and data landscape assessment
4. Feasibility conclusion and risk analysis (tied directly to the data)
5. Conclusions and references
6. Appendix (AI tool usage statement, if required)

Project 1: Optimizing High-Entropy Diboride Coatings

- Project Type: Optimization
- Data Source: AFLOW, Blade
- Inputs/features: Compositions
- Outputs/properties: E (max), Fracture Toughness (Max), Hardness (Max), Formation Energy (Min), Cost (Min), Thermal stability (constraint), Single Phase (constraint)

Project 2: Fusion Reactor 2nd Blanker Material

- Project Type: Predictive Tool
- Data Source: Find 3D Models and use OpenMC
- Inputs/features: Creep resistance, ductility, production of gas byproducts, material (constraint), more?
- Outputs/properties: structural (yield strength), efficiency (heat transfer and neutron transfer)

Project 3: MOSN Materials

- Project Type: Optimization
- Data Source: DFT Simulations
- Inputs/features: Composition of functional groups (2) and Position
- Outputs/properties: Band Gap close 1.3 (“Error”=|Band Gap - 1.3|, min), Elastic Modulus (max)

Project 4: Predict Solidification Microstructure

- Project Type: Predictive Tool
- Data Source: Thermocalc
- Inputs/features: Env Temp During solidification, Sample Geometry, Arc melter configuration (constraint), composition (constraint)
- Outputs/properties: 3D map of the sample (dendritic/planar)

Project 5: Eyewear Polymers: High Refractive Index for Thinner Lenses

- Project Type: Optimization
- Data Source: ?
- Inputs/features: budget (constraint?), material selection, geometry
- Outputs/properties: Thickness (min), mechanical durability (max), manufacturability (constraint), refractory index (constraint)

Project 6: Case Study to use LLMs to Accelerate Bayesian Optimization in High-Entropy Alloys

- Project Type: Optimization
- Data Source: Thermocalc, Borg
- Inputs/features: Composition
- Outputs/properties: Yield Strength (with LLM prior), Density, Pugh ratio, Phase (constraint)

Project 7: Carbon Fiber Structural Batteries

- Project Type: Optimization
- Data Source: COMSOL, Python Libraries (?)
- Inputs/features: 11 variables
- Outputs/properties: Ionic Conductivity (max), Internal Resistance (min), E (constraint), Shear Strength (constraint)

Today's Exercise

Create a "Master" Table

Example: Inputs, Inputs, Outputs, Constraints

ID	Co	Cr	Fe	Rec. Temp. (℃)	Holding time (h)	YS (MPa)	UTS/YS	Grain Size (μm)	Feasibility
XXX01	0.24	0.08	0.68	950	0.5	270.2	3.88	33.19	1
XXX02	0.16	0.24	0.6	1150	2	431.5	3.35	27.40	1
XXX03	0.24	0.24	0.52	850	4				0
XXX04	0.32	0.16	0.52	700	8	202.5	4.93	101.70	1