Graduation with Honors in Data Science

Every graduate of the College of Liberal Arts and Sciences with an upper-division grade-point average (GPA) of 3.50 or better receives at least a cum laude (honors) designation on his or her diploma; students not meeting this GPA criterion are ineligible for honors designations. ("Upper-division GPA" is defined as the GPA computed from all non-S/U courses taken at UF starting with the first semester in which the student enrolls after he/she has completed 60 hours. The courses themselves may be at any level.)

To be eligible for the designation *magna cum laude* (high honors), data science majors must write a thesis *in addition to meeting the GPA criterion of 3.50 or better.* To be eligible for the designation *summa cum laude* (highest honors), majors must write a thesis *and earn a GPA of 3.50 or better*. The Undergraduate Coordinator will make the final decision on the honors designation in consultation with the student's thesis supervisor and the Undergraduate Committee.

1 General Thesis Guidelines

The thesis must be neatly typed and formatted. The Graduate School's thesis style is recommended as a target for the quality of the format. The thesis should be grammatically correct and without spelling errors. The thesis must be mathematically correct, and must represent independent work by the student. Results need not be original, but the sources for unoriginal results must be clearly referenced. An abstract is required and it should begin with a definitive statement of the problem or project treated by the thesis. The purpose, scope, limits, and significance of the thesis should be clearly delineated, and the research methods, design, major findings, and conclusions should be described as concisely as possible.

The thesis must be written at a level understandable to the student's peers, i.e., to a strong senior student majoring in data science, unless the thesis is part of work to be published by a standard research journal aimed at experts in the field of data science .

Subject to the approval of the Undergraduate Coordinator, an article authored solely by the student may be submitted as the thesis if the article has been accepted by a standard research journal aimed at experts in the field of data science. In all other cases, the thesis should be written under the direct supervision of a faculty member from Departments of Statistics, Mathematics, or Computer & Information Science & Engineering. Faculty from other departments must first be approved by the Undergraduate Coordinator and can only co-supervise with a faculty member from one of the above three departments. The thesis

must be reviewed by the student's thesis supervisor for organization, content, grammar, and spelling before the initial submission. The student must adhere at all times to the University's Student Conduct and Honor Code

It is not permissible to submit the same thesis to multiple departments.

Past Honor's Theses can be located at the UF Undergraduate Honors Theses Collection.

Remark: Students who begin working with a faculty mentor early enough in their academic careers are encouraged to apply for the CLAS University Scholars Program. Application deadlines are usually in February, and only students who will graduate no earlier than May of the following calendar year are eligible to enter the program.

1.1 Length of thesis

It is up to the student's supervisor to determine an appropriate length for the thesis. Typical theses range in length from 15 to 30 pages. A theses may be shorter than this if it represents significant new research that is presentable in a shorter format. Expository theses may be longer. The length of the abstract should be 100-200 words.

1.2 Timeline and submission process

It is highly recommended that the student begin work on their thesis at least one semester prior to their final/graduating semester.

The student must agree to adhere to the following deadlines and procedures:

1. The Undergraduate Coordinator of the department in which the thesis is to be submitted shall be notified in writing (or email) of the intent to submit a thesis **between the first day of classes and no later than the last day of add/drop** of the graduating semester. Attached must be an abstract of the work to be submitted.
2. An Honors Thesis Submission Form must be submitted according to the instructions and deadlines found herein and UF Libraries Honors Theses under the "How do I submit?" tab.
3. A copy of the thesis, accepted by the supervisor as a final draft, must be submitted to the Undergraduate Coordinator no later than
   - March 30th for spring semester graduates,
   - July 6th for summer graduates, and
   - November 16th for fall semester graduates.
4. The Undergraduate Program Committee will then review the thesis and provide the student with a list of required and/or recommended revisions (if any). The student must complete all revisions and submit the revised thesis no later than
   - April 13th for spring semester graduates,
   - July 20th for summer graduates, and
   - November 30th for fall semester graduates.
5. A final version of the thesis must be submitted to the UF Library no later than the last

day of classes

- approx. April 24th for spring semester graduates,
- approx. August 9th for summer graduates, and
- approx. December 5st for fall semester graduates

2 THESIS GUIDELINES FOR THE DESIGNATION *Magna Cum Laude*

In order to earn the magna cum laude designation, an undergraduate thesis need not be a standard research journal-style data science paper, or work that could be rewritten as such a paper. Listed below are some acceptable general categories into which the thesis may fall. Other categories are possible, but the supervisor should confer with the Undergraduate Coordinator (who, in turn, may wish to confer with the Data Science Curriculum Committee) before directing the student into a category not listed below. Among the categories into which the thesis may fall are:

- **Applied modeling and analysis.** The student should apply appropriate data science methods to a real-world problem with nontrivial complexity. The analysis must go beyond basic applications, and the student should provide a clear rationale for methodological choices, discuss assumptions, and evaluate model performance. Discussion of ethical considerations (e.g., bias, fairness) should be included where relevant.
- **Data engineering or pipeline development.** The student should develop and document a working data pipeline (e.g., ETL processes, feature engineering, database integration) that supports large-scale or complex data analysis. The student should explain design choices and demonstrate reproducibility through clean, modular code repositories (e.g., GitHub).
- **Exploratory data analysis and visualization.** The student should conduct a sophisticated exploratory analysis of a rich dataset, identifying patterns, anomalies, and significant findings. The work must go beyond basic charts to include advanced, well-designed visualizations, interactive dashboards (if applicable), and clear interpretive commentary.
- **Software tool or reproducible research artifact.** The student should develop a working software tool (e.g., package, library, API, dashboard) or fully reproducible research project (e.g., Jupyter notebook, R Markdown report). The work should be properly documented and demonstrated to function correctly. If modeling is involved, the student should evaluate performance across different datasets or parameter settings.
- **Critical methodological review.** The student may produce a critical review of multiple data science methods applied to a common problem, assessing advantages, disadvantages, and real-world applicability. This must involve a practical application to data and a clear, well-organized comparative analysis, not just a literature review.

3 THESIS GUIDELINES FOR THE DESIGNATION *Summa Cum Laude*

In addition to the requirements for magna cum laude, the thesis should display originality in the solution of an acknowledged open scientific problem, exceptional mastery in apply-

ing, extending, or critically evaluating modern data science methods, demonstrated through complex real-world modeling, innovation in data engineering or modeling pipelines, responsible AI practices, or substantial, reproducible software artifacts.