what is the definition of integrity bias and fairness in software engineering?

In software engineering, particularly within the context of building ethical and responsible AI systems, "integrity," "bias," and "fairness" have specific definitions:

1. **Integrity**: This refers to maintaining a system's ethical and functional reliability. Integrity means building software that upholds moral and ethical standards, handles data securely, prevents unauthorized access, and respects user privacy. It also means ensuring that the software operates consistently and truthfully, without compromising on accuracy or distorting the intended outcomes.
2. **Bias**: Bias in software engineering refers to systematic favoritism or skewed treatment within a system that unfairly benefits one group over others. This often stems from unrepresentative data, historical biases in datasets, or algorithms that reinforce stereotypes. Bias can result in unequal or unfair outcomes and is usually categorized as:
    o **Data Bias**: Bias that arises due to an unbalanced dataset.
    o **Algorithmic Bias**: Bias introduced due to the algorithm's design or optimization goals.
    o **User Bias**: Bias that appears based on how users interact with or interpret the system.
3. **Fairness**: Fairness aims to ensure that software systems provide equal, unbiased outcomes for all user groups, particularly across sensitive attributes like gender, race, age, etc. Fairness in software engineering involves designing systems that actively work to reduce bias and unintended discrimination. Techniques to promote fairness include diverse data sampling, algorithm adjustments, and post-processing methods to adjust results and ensure parity across groups.

These principles are vital in software engineering, especially when developing systems that directly impact people, such as healthcare, hiring, and financial software.