

Vantage Glossary

Version 1.0

[Note: The following template is provided for use with the Rational Unified Process. Text enclosed in square brackets and displayed in blue italics (style=InfoBlue) is included to provide guidance to the author and should be deleted before publishing the document. A paragraph entered following this style will automatically be set to normal (style=Body Text).]

Revision History

Date	Version	Description	Author
11/10/2025	1.0		MAVENCODE team
04/07	1.1		MAVENCODE team

Table of Contents

	4
1. Definitions	4
1.1 <aTerm>	4
1.2 <anotherTerm>	4
1.3 <aGroupofTerms>	4
1.3.1 <aGroupTerm>	4
1.3.2 <anotherGroupTerm>	5
1.4 <aSecondGroupofTerms>	5
1.4.1 <yetAnotherGroupTerm>	5
1.4.2 <andAnotherGroupTerm>	5

Glossary

1.1 Access Points

Locations such as gates, roads, or paths identified during audio or video analysis, often relevant for land development assessments.

1.2 Actor

An external entity (user, drone, system) that interacts with the software to perform a use case.

1.3 AI Agent

A software program that can interact with its environment, collect data, and use that data to perform self-directed tasks that meet predetermined goals.

1.4 API (Application Programming Interface)

A programmatic interface (REST, gRPC, WebSocket) used to communicate with MCP subsystems. Allows different software components to communicate with each other.

1.5 API Gateway

The unified entry point for all multimodal services provided by the platform.

1.6 Artifact

Generated output from the system such as annotated videos, JSON files, or processed images that result from analysis tasks.

1.7 Audio Intelligence

Subsystem that analyzes recorded or streaming audio for sentiment, environmental cues, content extraction, and development insights.

1.8 Auto-Play Audio Mode

A configuration where text responses are automatically converted into speech output without explicit user request.

1.9 Bounding Box

A rectangular outline drawn around detected objects in images or video frames. Used in detection and tracking.

1.10 Business Objective

High-level goals the platform aims to achieve, such as high-performance multimodal analysis, accuracy targets, cost efficiency, customer impact, and market adoption.

1.11 Chat Interface

The user-facing interface through which users upload media, send commands, request analyses, or receive responses.

1.12 Classification Layer

A model output stage that categorizes detected objects into classes, including specialized or domain-specific ones.

1.13 Context Window

The maximum amount of text or multimodal data that an LLM can process in a single request. Overflows result in errors or warnings.

1.14 Coordinates (Destination Coordinates)

Geographic points provided by the user to navigate the drone to a specific location.

1.15 Depth Estimation

Process of determining spatial distance from the camera to objects in an image or video to support 3D

reconstruction.

1.16 Detection Pipeline

The system's workflow for identifying objects, features, and patterns in images or video frames.

1.17 Downsampling

An image processing technique that involves reducing the resolution or sampling rate of an image.

1.18 Drone

An unmanned aerial vehicle used to capture images, videos, audio, or execute navigation commands autonomously.

1.19 Drone Command

An executable instruction resulting from speech-to-text or manual coordinate input that controls drone navigation or behavior.

1.20 Endpoint (Navigation)

A target coordinate or location that the drone is instructed to reach. May be valid, unsafe, or inaccessible.

1.21 Event-Driven Requirement

An EARS requirement pattern where system behavior is triggered by an event such as a user uploading an image or giving coordinates.

1.22 FastAPI

The web framework used to implement MCP-compliant endpoints and services.

1.23 Feasibility Report

A structured summary generated from audio or video intelligence describing slope, vegetation, drainage, access, and constraints of a site.

1.24 Fine Tuning

A process similar to training, in which a user adjusts a model for specific tasks or domains.

1.25 Frame Processing

Running detection, tracking, or segmentation algorithms on individual video frames.

1.26 GPU (Graphics Processing Unit)

Hardware used for accelerating model inference for vision, audio, and multimodal tasks. Required for real-time performance.

1.27 GPU Acceleration

Use of graphics processing units to speed up model inference for vision, audio, and multimodal tasks. Required for real-time performance.

1.28 Hazard Zone / Unsafe Area

A location that the drone determines cannot be safely entered due to obstacles, environmental conditions, or inaccessible structures.

1.29 Image Preprocessing

Enhancement operations applied before analysis—denoising, cropping, resizing, or contrast adjustments—to improve model accuracy.

1.30 Inference (Model Inference)

Computational process where the model generates outputs such as captions, detections, or TTS audio.

1.31 Instance Segmentation

Pixel-level segmentation that identifies each object instance separately.

1.32 JSON Output

Structured metadata or analysis output generated from video or image processing (e.g., object detections and overlays).

1.33 Latency

The time required to process a request (image analysis, OCR, STT, etc.). Must meet strict performance targets (e.g., <200 ms).

1.34 LLM (Large Language Model)

Model used for text-based reasoning, visual Q&A, and multimodal processing. Provides answer generation and model switching.

1.35 MCP (Model Context Protocol)

A standardized communication protocol enabling all subsystems (speech, vision, spatial, tracking) to operate with shared context and structured inputs.

1.36 MCP Server

Model Control Protocol Server - handles routing and processing of requests between the user interface and AI models.

1.37 Model Switching

Ability to dynamically change vision, STT, TTS, or reasoning model backends depending on accuracy or speed needs.

1.38 Multimodal Input

Combination of image, video, audio, and text inputs processed in a unified pipeline.

1.39 Natural-Language Response

Model-generated textual description or answer in plain English returned to the user.

1.40 Navigation Route

A flight path computed by the drone to reach coordinates while avoiding obstacles and unsafe conditions.

1.41 Object Detection

Model capability that identifies items in images or video and returns bounding boxes with confidence scores.

1.42 Object Tracking

Following identified objects across multiple video frames to maintain continuity and ID assignments..

1.43 Panoptic Segmentation

A unified segmentation approach combining semantic and instance segmentation for detailed scene understanding.

1.44 Pathfinding

Computation of a safe navigation route using spatial awareness and segmentation.

1.45 Performance Metrics

Quantitative benchmarks such as accuracy, FPS, latency, concurrency, and memory usage used to assess subsystem performance.

1.46 Real-Time Processing

Model processing tasks (STT, video segmentation, object tracking) performed with low latency to support interactive use.

- 1.47 Routing Agent (MoE Router)
A system component that sends user requests to the correct MCP endpoint or model backend.
- 1.48 Safe Route
A drone navigation path that avoids obstacles, ensures integrity, and respects accessibility restrictions.
- 1.49 Segmentation (Semantic Segmentation)
Assigning each pixel in an image to a class label for detailed visual understanding.
- 1.50 Sensor Data
Audio, image, video, or spatial signals captured by drones or user devices for processing.
- 1.51 SLM (Small Language Model)
A language processing model designed to handle specific tasks with fewer parameters.
- 1.52 Spatial Reasoning / SpatialSense
Subsystem handling depth estimation, layout understanding, and 3D scene reconstruction.
- 1.53 Speech-to-Text (STT)
Model that transforms user audio commands or recorded sessions into structured text.
- 1.54 SVLM (Small Vision Language Model)
A model designed to handle vision and text modalities.
- 1.55 System Response
Any generated output—text, audio, JSON, or video—returned to the user after processing.
- 1.56 Temporal Consistency
Maintaining stable object identities and segmentation labels across video frames.
- 1.57 Text-to-Speech (TTS)
Model that converts textual responses into spoken audio output for hands-free use.
- 1.58 Timeout Condition
When processing takes longer than allowed, triggering warnings or fallback behavior.
- 1.59 Tracking Drop (ID Reassignment)
When an object's ID is lost across frames and the system assigns a new tracking ID.
- 1.60 Ubiquitous Requirement
EARS requirement that applies universally, with no trigger or condition.
- 1.61 Unsafe Condition
Environmental factor (obstacle, structure, hazard) that prevents safe navigation.
- 1.62 User
Any primary actor interacting with the system—real estate employees, analysts, surveyors, drone operators.
- 1.63 Vision Model
Any model used for image or video tasks such as captioning, detection, or segmentation. Users may switch between models.
- 1.64 Voice Command
Audio input by which a user directs drone behavior or queries the system. Converted via STT.

1.65 VoiceVision

Subsystem providing speech and audio intelligence: STT, TTS, diarization, emotion detection, and command recognition.

1.66 VQA (Visual Question Answering)

A technology that enables models to answer questions about visual content.

1.67 Agentic Intelligence

The capability of the platform to reason over user intent, select tools, and execute multi-step missions through LLM-driven agent orchestration.

1.68 Agentic Tool

A callable function, class, or MCP endpoint exposed to the LLM agent for performing a discrete capability such as drone control, image analysis, or TTS synthesis.

1.69 DroneController

1.70 The thread-safe Python class wrapping `djitellopy.Tello`, responsible for connecting to DJI Tello hardware, managing flight commands, and reporting telemetry.

1.71 FastMCP

A Python library used to build MCP-compliant tool servers and register AI tools with decorator syntax.

1.72 Gazebo

A robotics simulation engine used by VANTAGE to create the digital twin environment for safe drone testing.

1.73 MCP Client

The client-side component that connects to an MCP server and invokes registered tools from the mission agent or web service.

1.74 Ollama

A local LLM inference platform . It is used on VANTAGE to host language models such as llama3.2 for agentic reasoning.

1.75 ROS2

Robot Operating System version 2 middleware used for DDS communication, topic-based messaging, and integration between simulation and control components.

1.76 ROSBridge

A bridge that connects ROS2 topic messages to WebSocket clients, enabling frontend and cross-process communication.

1.77 WebSocket

A bidirectional communication protocol used by the web frontend and ROSBridge to exchange real-time telemetry, commands, and interface events.

1.78 Vue.js

The JavaScript frontend framework used to build the VANTAGE web UI.

1.79 Vite

The frontend build tool used for fast development, bundling, and hot reload in the VANTAGE web interface.