

ORCA

Democratizing music creation with AI-driven symbolic workflows

Orchestration and Recognition for Composition and Arrangement

Introduction



ORCA is a modular AI platform that unifies music composition, transcription, orchestration, and arrangement. It converts audio and sheet music into editable digital formats and supports symbolic transformations like transposition and genre-aware arrangement. Designed to enhance creativity, ORCA lowers barriers for independent artists and educators.

01

Introduction and Motivation

Music production tools and practices overview



Music production encompasses creation, recording, mixing, and mastering from ideation to final output.

Digital Audio Workstations (DAWs) like Logic Pro and Ableton enable sequencing and audio manipulation.

Notation software such as MuseScore and Finale supports symbolic editing and MIDI playback previews.

Definitions of transcription, arrangement, transposition, and digitization



Transcription converts audio to symbolic forms like MIDI or sheet music capturing pitch and rhythm.

Arrangement reworks pieces into new instrumentation while preserving core ideas.

Transposition shifts pitch to fit different instruments or vocal ranges.

Digitization converts analog inputs into machine-readable symbolic formats.

Limitations of current systems and generative AI trends

Current tools are fragmented, labor-intensive, and often support only monophonic transcription.

Generative AI usually uses a “prompt-in, product-out” paradigm, producing fixed audio with limited user control.

ORCA returns agency to users via tagging-based symbolic generation, enabling editable outputs and creative workflows.



02

ORCA System Components

Optical Music Recognition (OMR) system



Uses domain-pretrained encoders and a 2D cross-attention decoder for robust sheet music transcription.

Retains spatial layout to improve pitch and accidental detection, converting images into MusicXML.

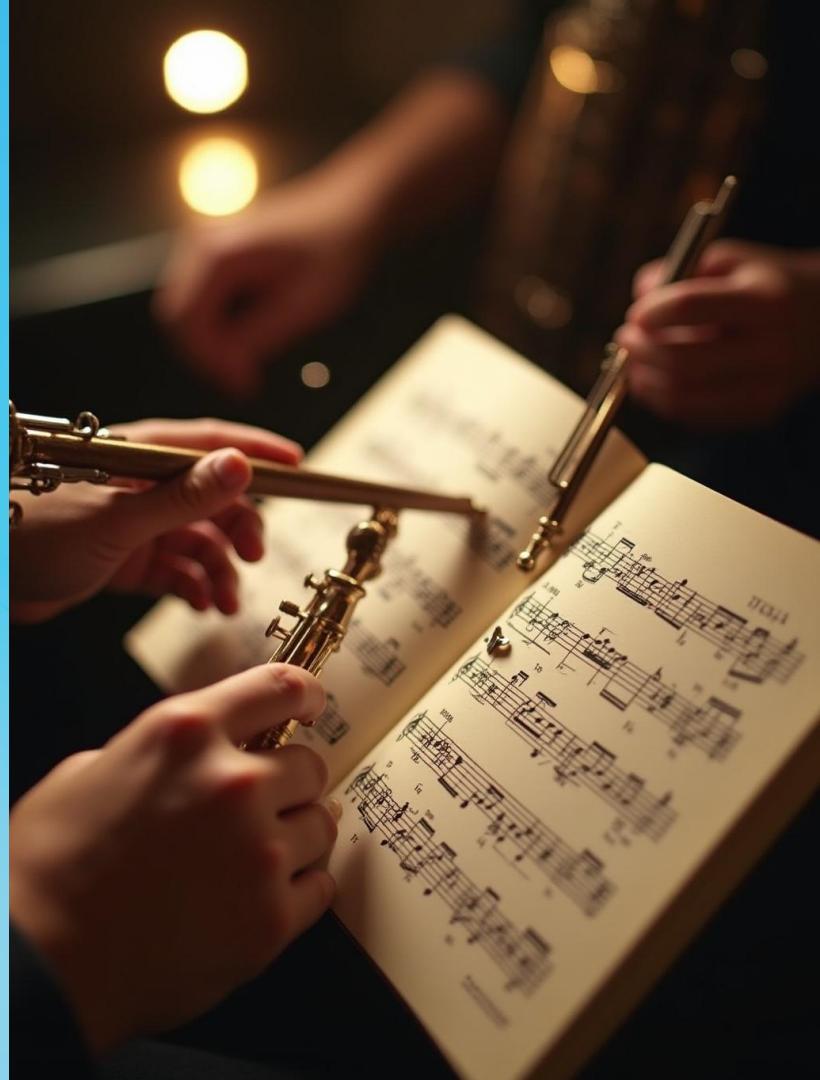
Supports complex multi-system orchestral layouts, forming the basis for transcription correction and arrangement.

Instrument Transposition (IT) system

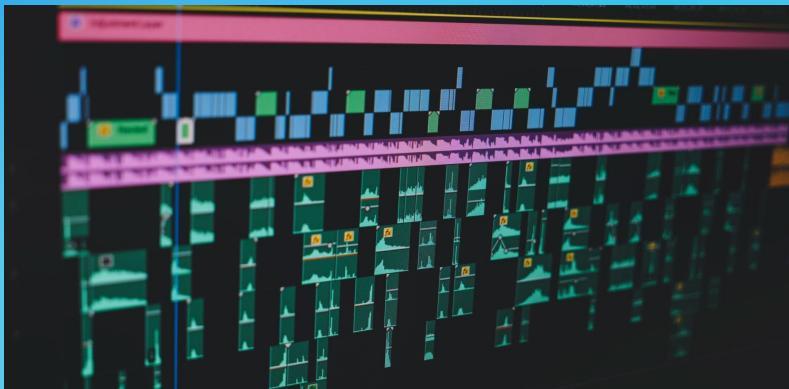
Transforms music symbolically between instruments respecting pitch range, tessitura, articulation, and dynamics constraints.

Uses hybrid rule-based and transformer-based sequence models for idiomatic phrasing and style preservation.

Enables real-time auditioning and rendering, adapting melodies to different instruments while preserving musical intent.



Automated Music Arrangement (AMA) system

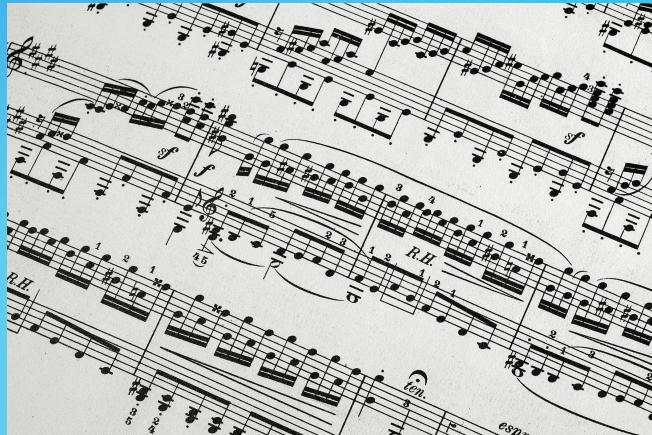


Performs data-driven orchestration assigning instruments to symbolic scores with style and structural coherence.

Models texture, layering, and genre-aware orchestration using Transformer-based sequence-to-sequence learning.

Supports composer augmentation with 86% structural similarity to professional arrangements and high user ratings.

Limitations of current music production systems and generative AI trends



Current systems suffer from **limited expressivity** in audio rendering; nuances like microtiming and vibrato are not fully captured.

Training data biases favor Western tonal music, limiting **genre and instrument coverage**. Evaluation metrics don't fully reflect musical quality.

Most tools lack **real-time interactivity**, operating offline and restricting live or collaborative use.

ORCA aims to overcome these through reinforcement learning, multimodal integration, and performance style conditioning.

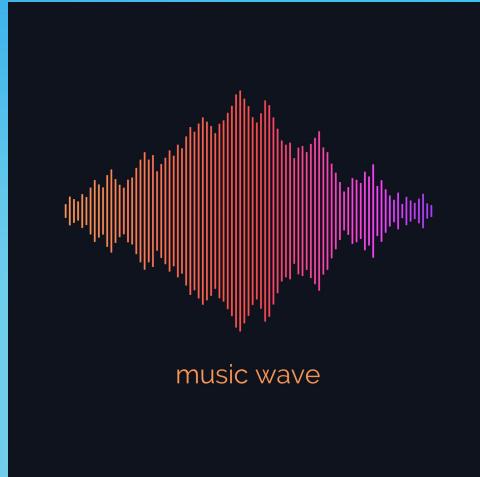
03

ORCA System

Architecture and

Components

Optical Music Recognition (OMR) with domain-pretrained vision encoders and 2D cross-attention decoder



OMR converts scanned sheet music into editable **MusicXML** using ConvNeXt or Swin Transformer encoders **pretrained on large sheet music datasets**.

A 2D cross-attention decoder preserves spatial relationships, improving **pitch alignment** and **accidental detection**. Auxiliary tasks like staffline segmentation enhance accuracy.

Supports complex multi-system orchestral layouts enabling direct symbolic editing and further processing.

Instrument Transposition (IT) as constrained sequence transformation with hybrid rule-based and neural models

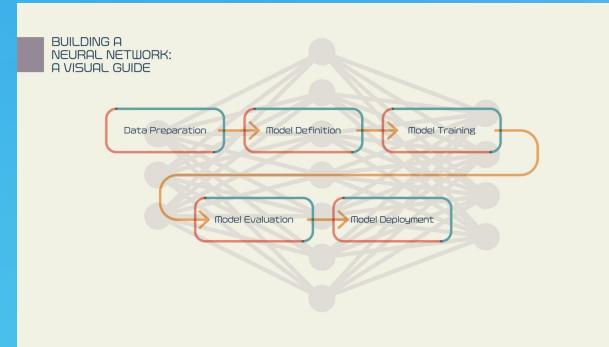
IT adapts symbolic music between instruments respecting pitch range, tessitura, articulation, and dynamics constraints.

Combines **rule-based filtering** to enforce feasibility with a Transformer-based neural model for idiomatic **phrasing** and **expressive mapping**.

Enables real-time audition and seamless integration, transforming polyphonic piano textures into suitable formats for monophonic instruments.



Automated Music Arrangement (AMA) using Transformer-based sequence-to-sequence modeling conditioned on genre and instrumentation



AMA produces multi-instrument arrangements from single-track MIDI, modeling texture, layering, and timbral continuity. A Transformer encoder-decoder is trained on the Lakh MIDI dataset and conditioned on genre embeddings, producing stylistically coherent orchestrations. Post-processing smooths dynamics and consolidates parts, outputting fully orchestrated MIDI or MusicXML for editing, with 86% structural similarity to professional arrangements.

Conclusions



ORCA presents a **unified symbolic-first AI platform** that democratizes music creation by blending transcription, transposition, arrangement, and generation. Its modular pipeline, grounded in MIDI, ensures **interoperability, editability, and modular improvement**.

By emphasizing **human-in-the-loop workflows**, ORCA empowers creatives of varied expertise to produce professional-grade music and fosters inclusivity across cultural and educational contexts.

Future improvements target enhanced expressivity, expanded genre coverage, and real-time interactivity, advancing symbolic music intelligence systems.

Current systems lack full **expressive playback**, missing nuances like microtiming and vibrato. Training data mainly covers Western tonal music, limiting **genre and instrument diversity**. Evaluation metrics do not fully capture **musical quality**, and systems mostly work offline without **real-time interaction**.

ORCA addresses these with reinforcement learning, multimodal data integration, and style conditioning.