



Tutorial

**VENICE
2025**

Theme

Business Models for SME in the AI Era

IARIA Congress 2025 & DigiTech 2025



Meet Petre

VENICE
2025

Lecturer

Prof. Dr. Petre Dini, IARIA, USA/EU



PETRE – AI-related activities

VALENCIA
November 2023



- Petre: 1980/90
 - Fuzzy-based resource allocation, Automatic knowledge incorporation, CAD/CAM Expert Systems,
 - Real-time embedded systems, Space/time thinking and processing, Multi-layers context-based meaning
- Petre: 1992: The First ITC Conference (Montreal), tutoring systems, self-adaptable Q&A professor-student systems (advanced Chatbots)
- Petre: 1997 Dartmouth, Mobile Intelligent Agents (Intelligent Grasshopping Polling)
- Petre: 1997-2000: Nomadic code, Mobile agents, (Grasshopper EU project)
- Petre: 2000-2010: Autonomous systems, Policy-driven systems, Intelligent systems (pushed to Patents, ITU, TMF, standards)
 - Capturing emerging properties, Variable pooling frequency, Self-adaptable decision policies, Reflexive-policies (Digital-Twins)
 - Routers embedded-AI (temporal logic in Syslog processing, policy-driven signal processing)
- AI-driven Selection of Content Servers based on Current Server Availability (dynamic availability, heuristics, real-time)
- Petre: 2010 - now (active observer and critic, panels, open discussions)



Petre DINI
petre@iaria.org



Themes

VENICE
2025

Modus operandi in the pre-AI era

Challenges of classical models

Modus operandi using AI-based tools/platforms/approaches

Challenges in AI era; platform/tools (LLMs) harmonization

AI-related corporation skills; agility, licensing

AI-related employee skills; knowledge engineer



Modus operandi in the pre-AI era

VENICE
2025

Decision-Making Style Operational Schema Technology Use Customer Engagement Workforce and Skills

Relationship-based loyalty: Success depended on trust, proximity, and personal service rather than algorithmic targeting.

Word-of-mouth marketing: Branding was local and reputation-driven, with low-scale advertising.

Generalist staff: Employees often wore multiple hats with broad but shallow skill sets.

Low digital literacy: Training was ad hoc, and digital tools were only used when necessary.

Experience-driven: Decisions based on intuition, owner/founder expertise, and historical performance.

Flat hierarchy: Quick decision loops with minimal bureaucratic overhead (centralized in the hands of a small leadership group).

Reactive (vs predictive) business model: Response to market signals and crises were rather defensive.

Manual and human-centric processes:

Customer relations handled via direct contact, phone, or email.

Inventory, sales, and financial records often managed in spreadsheets or paper-based ledgers.

Departmental silos: Functions like marketing, sales, and operations were loosely integrated

Low data dependency: Limited data collection, minimal analytics—decisions were made with incomplete or aggregated data.

Basic IT tools:

Use of desktop software like Microsoft Office (Excel, Word, Access).

Basic accounting and point-of-sale systems.

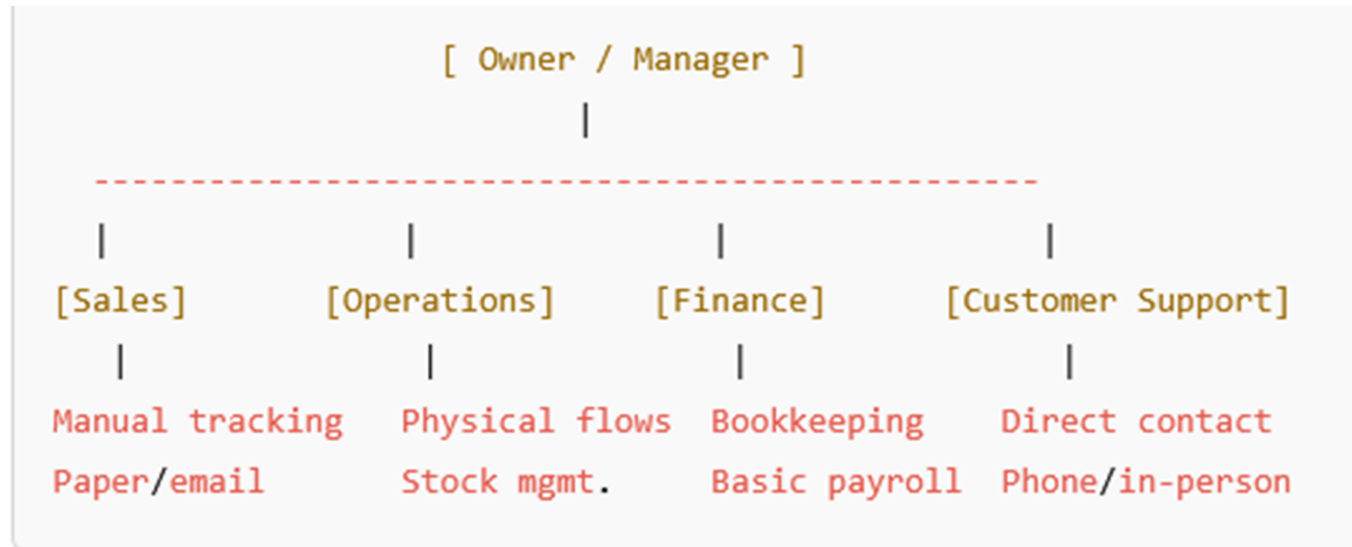
Minimal automation: Repetitive tasks were done manually; automation was costly or inaccessible.

Limited digital presence: Many SMEs had static websites, if any, with low adoption of e-commerce or digital marketing.



Schema and Challenges

VENICE
2025



Support Tools

MS Office Suite

On-premise hardware (printers, servers)

Local storage (files, backups)

Data Flow

Minimal, unstructured, delayed (days/weeks)

Challenges

Investors: Single point-of-failure syndrome

Strategy

Marketing

Business continuity

Market predictions

Innovation

Personnel retention

AI-concepts?

Flexible approach

Q: Risky



Pre-AI vs Digital/Post-AI

VENICE
2025

Dimension

Decision-Making Experience-based, intuition-driven, reactive || Data-informed, predictive, AI-assisted (dashboards, recommender systems)

Business Strategy Survival/growth based on local niche, slow adaptation || Agile, data-driven models, leveraging AI insights for scalability

Process Management Manual, spreadsheet- or paper-based tracking || Workflow automation, robotic process automation (RPA), smart process orchestration

Customer Engagement Personal interaction, word-of-mouth loyalty || AI-powered chatbots, personalization engines, CRM with sentiment analysis

Marketing & Sales Static website, local ads, generalist campaigns || SEO/SEM, social media analytics, targeted ads via ML models
Product/Service Design Based on owner vision or past feedback || Real-time feedback analysis, A/B testing, demand prediction via AI

IT Infrastructure On-premise systems, minimal integration || Cloud-native platforms, APIs, scalable digital ecosystems

Data Handling Little to no structured data use || Real-time data pipelines, analytics dashboards, predictive models

Financial Management Bookkeeping via spreadsheets or basic tools || AI-based forecasting, fraud detection, dynamic pricing models

HR & Workforce Skills Generalists with manual task focus || Roles evolving to human-AI collaboration, upskilled digital-savvy staff

Supply Chain Local suppliers, limited traceability || AI-optimized logistics, dynamic sourcing, real-time supply chain monitoring

Innovation Capability Low (due to cost and risk aversion) || Open innovation, AI for prototyping, fast feedback loops

Security & Compliance Manual checks, low-level awareness || AI-driven anomaly detection, compliance automatio



... the outcome

**VENICE
2025**

Loosing control/ownership

On-demand tools (license, versioning, product ownership)

Need on narrow (and quickly evolving/changing) skills

Quick deskilling

Highly dependent on critical resources (human and digital)

Offer a tailored product/service built on top of proven components, with added value in customization, domain expertise, or integration.

 **Infrastructure/Tools**

Licensed APIs (e.g., OpenAI, Anthropic), FOSS under LGPL/MIT, or open-core systems

 **Intelligence Layer**

LLMs, vector DBs, autoML platforms licensed per-use or hosted

 **Proprietary Layer**






SME-developed: workflow orchestration, UX, domain-specific logic, rule-based modules



Revenue Stream and Licensing





VENICE
2025

Revenue streams

-  **B2B Licensing** License the system as a SaaS or on-prem platform (monthly/annual tiers)
-  **Customization Services** Custom workflows or modules tailored to client processes
-  **API as a Service** Expose SME-layer functionality via custom APIs
-  **Integration Add-ons** Plugins for CRMs, ERPs, sector-specific software (education, law, medical)
-  **Analytics / Reporting** Premium dashboards or compliance modules
- CRM** Customer Relation Management
- ERP** Enterprise Resource Planning

Licensing strategy

Use tools under licenses that allow derivative/commercial work or that are “as-a-service” (API-based)

Tool Type	Example	SME Strategy
 LLMs	OpenAI, Claude, Cohere	Use via API – no model redistribution
 Open-source libs	spaCy, HuggingFace, FastAPI	Comply with licenses (MIT/Apache = safe)
 Open-core	LangChain, Supabase	Use core, pay for commercial support if needed
 Cloud APIs	AWS Comprehend, Azure ML	Usage-based costs; no IP lock-in

Be careful with:

GPL code (may require open-sourcing your product)

Redistributable binaries if licensing terms are restrictive

GNU GPL: General Public License (GPL)

TP-LINK GPL code:

Third Party : partly contain software code developed by third parties, including software code subject to the GNU General Public Licence (“GPL”), Version 1/Version 2/Version 3 or GNU Lesser General Public License(“LGPL”)



Ownership Map / Benefits

VENICE
2025

Ownership

Component

User interface, workflow logic
Prompt engineering / orchestration
Models (if via API)
Conceptual IP (how system works)

Ownership

Fully SME-owned
SME-defined
Not owned – licensed
Ownable + protectable

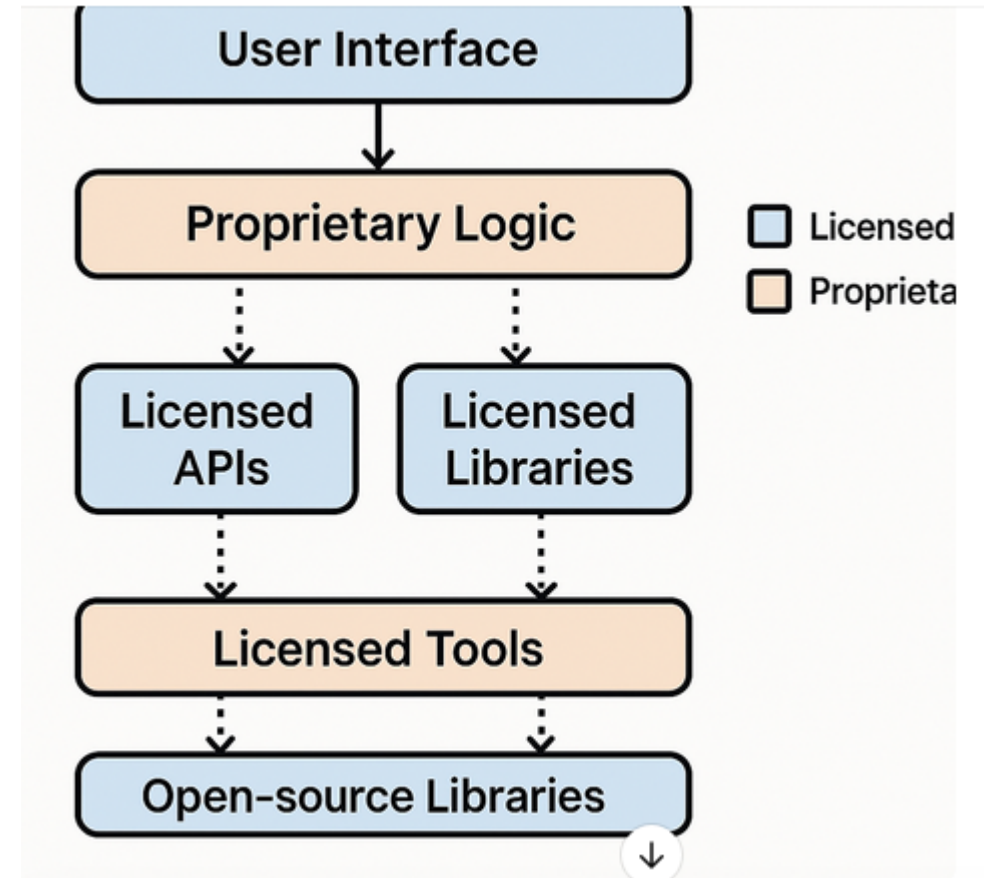
Strategic Benefits

Benefit

- 🎯 Focus on core value
- ⌚ Faster time-to-market
- 💰 Controlled costs
- 🔑 Maintain IP
- 🔄 Upgradeable backend

Outcome

SME builds what matters most: domain logic, UX, compliance
Reuse of proven tools reduces dev overhead
Pay-as-you-grow API or license usage
SME owns workflow, brand, integration layer
Swap tools as better ones emerge

















Roles / Skills / BU Size

VENICE
2025

Roles

Role	Responsibility	Team Size
 AI Architect / Engineer	Designs how to integrate APIs, models, and internal logic	1–2
 Prompt/Task Designer	Engineers LLM prompts, evaluates model behavior, controls cost & quality	1
 Domain Expert / Product Owner	Defines real user needs, validates utility, tracks licensing constraints	1
 Full-Stack Developer	Builds the interface, workflow, and connections to libraries/APIs	1–2
 Data/Compliance Advisor (part-time or external)	Ensures licensing/legal/safety boundaries are respected	Optional/external

Core Skills Matrix

Skill Area	Concrete Skills	Why It's Needed
 Model Awareness	Know top LLMs (GPT-4, Claude, Cohere, Mistral, LLaMA), licensing, strengths, limits	To select and justify model usage, e.g., open vs API vs local)
 Prompt Engineering	Few-shot prompts, chain-of-thought, reasoning control, role setting	To shape behavior and optimize API efficiency
 Tool Integration	Use of LangChain / LlamaIndex / OpenAI API / HuggingFace / vector DBs	To glue tools together into working pipelines
 Basic Concept Modeling	Concept extraction, graph building (e.g., Neo4j, RDF-lite)	If adding a concept layer (LCM), for structure and reasoning
 Evaluation Design	Know how to assess output quality (fluency, truth, domain fit)	So your system doesn't hallucinate or mislead users
 Licensing/Legal	Understand commercial rights of APIs, open source, data policies	To avoid IP or compliance risk
 Human-Centered Thinking	Lean UX principles, feedback loops	To validate product-market fit early and cheaply



Effectiveness / Product Team

VENICE
2025

How to Stay Lean Yet Effective

Strategy

✂️ Use modular open-core

embeddings, chains, vector stores — use LangChain, Supabase, etc.

⚖️ Start with one model family

Choose GPT or Claude as your “core” and test others only when needed

📋 Reuse templates and evaluators

Use prompt collections, eval chains (e.g., Ragas, Trulens)

🔄 Human-in-the-loop

Design feedback capture early (annotation, correction, thumbs-up/down)

👤 Pair with strong domain expert

Even better than tuning models is tuning the workflow logic and vocabularies

Description

Don't reinvent

Choose GPT or

Claude as your “core” and test others only when needed

Design feedback

Even better than

tuning models is tuning the workflow logic and

vocabularies

Applied AI Product Team

Core Elements

AI-Driven Solutions
Domain Expertise
User-Centric

Core Skills

- Model Awareness
- Prompt Engineering
- Tool Integration
- Basic Concept Modeling
- Evaluation Design
- Licensing/Legal

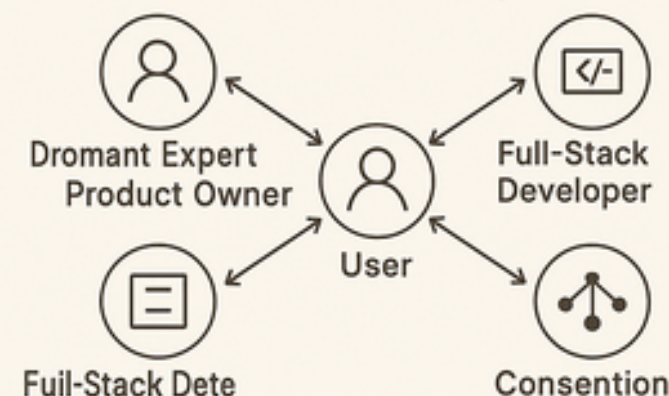
Key Tools

- LangChain
- LlamaIndex
- OpenAI API
- Vector DBs

Key Roles

- ⚙️ AI Architect / Engineer
Integrate LLMs, APIs, libraries
- 🗣️ Prompt Designer
Create effective prompts, refine outputs
- 👤 Domain Expert / Product Owner
Define requirements, validate solutions

Interaction Map










Roles / Tools / Licensing

VENICE
2025

Team Roles & Assignments

Role	Name / Notes	Time Allocation
 AI Architect	Engineer	Full-time / Part-time
 Prompt Designer	Evaluator	
 Domain Expert	Product Owner	
 Full-stack Developer		
 Legal	Compliance Advisor (optional/external)	

Core Tools & Frameworks

Tool Type	Selected Tool	Why This Tool?
LLM API	(e.g., OpenAI GPT-4, Claude, Mistral)	
Concept Engine	Graph	(e.g., Neo4j, RDF)
Prompt Chains	Lang	(e.g., LangChain, CrewAI)
Embedding Store	(e.g., Chroma, Pinecone, FAISS)	
UI / App Framework	(e.g., React, Streamlit, FastAPI)	

Iteration & Evaluation Plan

Evaluation Focus	Approach / Metric	Frequency
Output Quality	Human eval + metrics (BLEU/Faithfulness)	Weekly
User Feedback	Real-user pilot, scorecards	Monthly
Cost Optimization	Token usage monitoring	Every sprint

Risk & Licensing Checklist

Area	Action Required	
Status		
API Terms	Review commercial use terms	<input checked="" type="checkbox"/>
Data Privacy	Compliant with GDPR / local regs	<input type="checkbox"/>
Open Source	Track LGPL/GPL obligations	<input type="checkbox"/>
Prompt/IP Boundaries	Avoid sensitive reuse	<input type="checkbox"/>



Summing-up: Knowledge Engineer

VENICE
2025

Note: a paradigm shift where traditional roles (e.g., software architect, QA analyst, DB admin) are morphing or being recomposed into fluid, cross-functional profiles in response to:

 LLM integration

 Tool orchestration instead of ground-up building

 Semi-autonomous agents replacing hard-coded pipelines

 New licensing/compliance terrains

 Cognitive layer design (prompt logic, concept graphs, task decomposition)

A “developer” today might need to think like a knowledge engineer, while a “product owner” now needs to understand model reasoning boundaries.



Q&A

VENICE
2025

Q

FYI (on licensing):

<https://developer.nvidia.com/blog/curating-trillion-token-datasets-introducing-nemo-data-curator/>
Curating Trillion-Token Datasets: Introducing NVIDIA NeMo Data Curator

