PULSE Agent - Technical Research

Date: 2025-11-19 **Purpose:** Technical research for Role #1 (PULSE - Lead Finding Agent) **Workshop Reference:** Session 1 notes (20251119-workshop-session-1.md)

Role #1: PULSE Agent Overview

From Workshop Notes:

- **Purpose:** Find leads by identifying market needs/pain points
- **Method:** Web scraping and crawling for keywords
- **Execution:** Periodic runs (hourly/12hr/24hr - TBD)
- **Key Example Triggers:**
    - Heavy activity around "ChatGPT is not sufficient"
    - Articles about "massive shortage of translators in Germany"
- **Target:** Access company registration databases (e.g., Germany)
- **Output:** Pull relevant client contacts/emails, send outreach

---

Lead Scraping Tools & APIs (2025)

Top Platforms for Lead Generation

1. **Apify** (Recommended for Production)

- **Strengths:**
    - Enterprise-grade with API support
    - Export to CRM integrations
    - Schedule automated runs
    - Specific Handelsregister API for German companies
- **Use Case:** Large-scale, scheduled scraping with CRM integration
- **Pricing:** Variable (usage-based)
- **API:** Full REST API support

2. **Browse AI**

- **Strengths:**
    - Transforms websites into custom API endpoints
    - No-code setup
    - REST API for direct integration
- **Use Case:** Quick API endpoint creation from any website
- **Integration:** Direct API calls from PULSE agent

3. **Scrapy** (Open Source)

- **Strengths:**
    - Free and open-source
    - Python framework (easy integration)

        - Built for large-scale projects
- **Use Case:** Custom implementation with full control
- **Pricing:** Free
- **Learning Curve:** Requires Python development

4. **PhantomBuster**

- **Strengths:**
    - Pre-built scrapers for LinkedIn, Twitter, etc.
    - Social media lead generation focus
- **Use Case:** Finding leads from social platforms

5. **AI-Powered Options**

- **AIScraper:** Chrome extension with AI-powered extraction
- **Thunderbit:** 2-click scraping with AI
- **Benefits:** Minimal setup, automatic data extraction using ML/NLP

No-Code Integration Options

- **Make, n8n, Zapier:** Can integrate APIs using API keys
- **Benefit:** Rapid prototyping without coding

German Company Registration APIs

Critical for PULSE Agent (Janus specified Germany example)

1. **handelsregister.ai** (Recommended)

- **Access:** Entire German commercial register via REST API
- **Update Frequency:** Daily
- **Data:** Structured, reliable, fast
- **Key Features:**
    - Company details and contacts
    - Management/officers information
    - Real-time access
- **URL:** https://handelsregister.ai/en

2. **OpenAPI - Company Advanced DE**

- **Data Points:** 60+ real-time data points
- **Includes:**
    - Company details
    - Financial data
    - Contact information (website, phone)
    - Business classification
- **Tiers:** Company Start DE (basic), Company Advanced DE (full contacts)

3. **Implisense (via RapidAPI)**

- **Strengths:**
  - Credible, checked, cleaned data from Handelsregister
  - Daily updates
  - Advanced search/filtering:
    - Product names, technologies, certifications
    - Location, industry, size filters
- **Use Case:** Finding specific company types (e.g., companies needing translation services)

4. **OpenRegisters Handelsregister API**

- **Features:**
  - Real-time access
  - Company addresses and contacts
  - Management and authorized officers
  - Capital structure
- **Reliability:** Structured and fast

5. **Apify Handelsregister API**

- **Specialization:** Extracts from handelsregister.de directly
- **Data:**
  - Company data
  - Shareholders
  - Addresses
  - Register court and documents

Technical Architecture Recommendations

Phase 1: Proof of Concept (Week 1-2)

```
1. Use Browse AI or Thunderbit for quick prototyping
   - Set up monitors for keywords like:
     - "ChatGPT limitations"
     - "AI not sufficient"
     - "shortage of [service]"
   - Test webhook/API delivery
2. Test handelsregister.ai API
   - Sample 10-20 companies
   - Verify contact data quality
   - Check rate limits and pricing
```

Phase 2: MVP Implementation (Week 3-4)

```
1. Build PULSE Agent Core:
```

```
   - Python-based scraper (Scrapy or custom)
   - Keyword detection logic
   - Spike/cluster identification algorithm
2. Integrate German Company API:
   - Choose: handelsregister.ai or Implisense
   - Build company lookup function
   - Extract contact emails/info
3. Outreach Module:
   - Email template system
   - "Try Meet Mike" messaging
   - Website redirect tracking
```

Phase 3: Production (Week 5-6)

```
1. Scheduling System:
   - Hourly keyword scans (news, forums, social)
   - 12hr deep analysis (trend identification)
   - 24hr company database refresh
2. Intelligence Layer:
   - Pattern recognition (what problems are trending?)
   - Relevance scoring (how well does Meet Mike solve this?)
   - Priority ranking (which leads to contact first?)
3. Compliance & Ethics:
   - GDPR compliance (German data protection)
   - Opt-out mechanisms
   - Rate limiting (don't spam)
```

Key Technical Considerations

1. Data Sources for Keyword Detection

Where to scrape for "desperate clients":

- News aggregators (Google News API, NewsAPI.org)
- Reddit (via API) - subreddits for business problems
- LinkedIn (PhantomBuster) - job postings indicating needs
- Industry forums and Q&A sites
- Twitter/X (API) - trending business complaints

2. Spike/Cluster Detection Algorithm

Approach:

```
# Pseudocode for spike detection
def detect_market_spike(keyword, timeframe='24h'):
    baseline = get_historical_average(keyword, days=30)
    current = get_current_mentions(keyword, timeframe)
    if current > baseline * 2.0:  # 2x increase = spike
        return {
            'spike': True,
            'magnitude': current / baseline,
            'keyword': keyword,
```

```
            'relevance': calculate_relevance_to_meet_mike(keyword)
        }
```

3. Email Validation & Deliverability

Tools:

- NeverBounce API (email verification)
- ZeroBounce (validation + spam check)
- Hunter.io (find company emails)

4. GDPR Compliance (Critical for Germany)

Requirements:

- Legitimate interest or consent for contact
- Easy opt-out in all communications
- Data retention limits (delete after X days if no response)
- Privacy policy transparency
- **Recommendation:** Consult with Christian on legal requirements

5. Rate Limiting & API Costs

Budgeting:

- API calls: Track costs per lead found
- Email sends: Consider cost per outreach
- Database queries: Bulk vs. individual lookup pricing
- **Estimate:** Budget $200-500/month for initial testing

Integration with Meet Mike Ecosystem

Workflow: Lead Discovery · Contact · Website

```
1. PULSE detects spike: "Companies complaining about AI translation
quality"
2. PULSE queries Handelsregister API:
   - Filter: Translation services companies in Germany
   - Size: 50-500 employees (sweet spot)
   - Extract: Contact emails, decision makers
3. PULSE sends outreach:
   Subject: "We noticed challenges with AI translation - Meet Mike
can help"
   Body: "We've developed a solution specifically for [their
problem]..."
   CTA: Visit website · Next agent (Role #2) takes over
4. Role #2 (Website Engagement) continues the journey...
```

Handoff Protocol Between Agents

- PULSE · Role #2: Pass lead context (problem detected, company info, interest level)

- Role #2 · Role #3: Pass conversation history, qualification status
- Each agent maintains context for seamless client journey

Immediate Next Steps for Workshop

When Janus resumes:

1. **Validate Assumptions:**
   - Confirm: Is Germany the primary market or just an example?
   - Confirm: What types of problems does Meet Mike solve best?
   - Define: What's the ideal client profile (company size, industry)?
2. **Technical Decisions:**
   - Choose scraping platform (recommend: Apify for production, Browse AI for MVP)
   - Choose German API (recommend: handelsregister.ai for reliability)
   - Define: Hourly vs. 12hr vs. 24hr run frequency
3. **Development Blockers to Address:**
   - Christian's input needed on:
     - GDPR compliance approach
     - Budget allocation for API costs
     - Legal review of automated outreach
   - DON'T START coding until these are resolved (per workshop notes)
4. **Define Remaining Roles (2-6):**
   - Role #2: Website engagement (chatbot? form? video?)
   - Role #3: Price negotiation (automated? human-assisted?)
   - Role #4: Problem/solution handling (technical scoping?)
   - Role #5-6: (To be defined by Janus)

Resources & References

APIs:

- Handelsregister.ai: https://handelsregister.ai/en
- Apify: https://apify.com
- Browse AI: https://www.browse.ai
- Implisense: https://rapidapi.com/Implisense/api/german-company-data

Tools:

- Scrapy (Python): https://scrapy.org
- PhantomBuster: https://phantombuster.com
- Make/n8n/Zapier: For no-code integration

Reading:

- Web Scraping for Lead Generation: https://www.smartlead.ai/blog/web-scraping-for-lead-generation
- GDPR Compliance Guide: (Christian to provide legal resources)

Status

Research Status: · Complete **Ready for:** Workshop Session 2 when Janus resumes **Dependencies:** Christian's input on GDPR/legal, budget approval **Next:** Define Roles #2-6 in workshop, then begin technical planning

---

Prepared by: Claude (Team Lead) **Date:** 2025-11-19 20:15 UTC **Context:** Autonomous work - preparing materials for paused workshop