PROYECTO INTEGRADOR II

ACTA DE CONSTITUCIÓN DEL PROYECTO INTEGRADOR II 30-01-2025

Juan Andres Contreras Hernandez

Karen Nicol Dayanna Lizarazo Ortiz

Raul Ferney Lozano Navarro

Maria Angelica Parra Grazt

Jerson Arley Porras Lozano

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Facultad de Ingeniería de Sistemas e Informática

Universidad Pontificia Bolivariana Seccional Bucaramanga

Bucaramanga,30 de enero de 2025

[ACTA DE CONSTITUCIÓN DEL PROYECTO INTEGRADOR (PROJECT CHARPET) 4](#_heading=h.rzoti3jtf7fo)

[1. Información general del Proyecto Integrador 4](#_heading=h.b3bkj99o29em)

[1.1. Nombre del Proyecto Integrador 4](#_heading=h.t1trjvcmvc0p)

[1.2. Promotor del proyecto Integrador 4](#_heading=h.idzgby2qtyu3)

[1.3. Patrocinador del Proyecto Integrador 4](#_heading=h.2lj5d1rq087l)

[1.4. Definición del Product Owner1.5. Definición del Scrum Master 4](#_heading=h.328dyb44xk7v)

[1.6. Definición del Scrum Team 4](#_heading=h.pgci0wn1xxl8)

[1.7. Interesados (Stakeholders) 4](#_heading=h.h839xp86i5eh)

[1.8. Socio(s) 5](#_heading=h.ae9ncmxgey1l)

[1.9. Fecha de inicio esperada 5](#_heading=h.gf7kx9e1hhh6)

[1.10. Fecha de finalización 5](#_heading=h.w8zd0boyh10n)

[2. Detalles del Proyecto Integrador 5](#_heading=h.3rix9q73xenp)

[2.1. Declaración de la visión del Proyecto Integrador 5](#_heading=h.3x6a8kfvnxvg)

[2.2. Planteamiento del problema 5](#_heading=h.uuucb4qlrzao)

[2.3. Misión 6](#_heading=h.ym6zowlyqxb8)

[2.4. Visión 6](#_heading=h.df4sx3fv5yi0)

[2.5. Alcance 6](#_heading=h.mp27ddluefc8)

[2.6. Objetivos 7](#_heading=h.t5o75xdos32s)

[2.7. Beneficios esperados 9](#_heading=h.zamlhhpsn6fm)

[3. Especificación de requerimientos 9](#_heading=h.t39r0l33li67)

[Requerimientos Funcionales: 9](#_heading=h.wmwbilbqgaah)

[Requerimientos No Funcionales: 9](#_heading=h.sxvop5656y6)

[Requerimientos Técnicos: 10](#_heading=h.n9emhm2tctcy)

[4. Identificación de riesgos 10](#_heading=h.dkev1e6hhlb2)

[5. Criterios de aceptación del Proyecto Integrador 11](#_heading=h.a6r0tooaimop)

[Criterios Funcionales: 11](#_heading=h.16j1jjxgg3g7)

[Criterios Técnicos: 12](#_heading=h.2omarlf1xfot)

[Criterios de Calidad: 12](#_heading=h.uv9nc0clvmha)

[5.1. Duración del Sprint 12](#_heading=h.nevoeh8qsp4a)

[5.2. Criterios de terminado (Done) 12](#_heading=h.2o5m1x3764x8)

[6. Cronograma de lanzamientos 13](#_heading=h.pn2q7sg4vgxh)

[Hitos Importantes: 13](#_heading=h.gdrufv5agsre)

[7. Documentación de soporte 14](#_heading=h.ci8m23gj68zc)

[Documentación Técnica: 14](#_heading=h.h80h0wm93x29)

[Documentación de Usuario: 14](#_heading=h.aonilnnojuqq)

[Documentación de Proyecto: 14](#_heading=h.k2xcx9c8hjdq)

[8. Recursos Requeridos 15](#_heading=h.mh3wl3i340ba)

[8.1 Recursos Humanos 15](#_heading=h.u7ozo8u0t0yr)

[8.2 Recursos Tecnológicos 16](#_heading=h.5277t0td9958)

[9. Estructura de Desglose del Trabajo (EDT) 16](#_heading=h.8eapqzldb15o)

[Fase I: Fundación (Semanas 1-2) 16](#_heading=h.11mo5cz7ze6z)

[Fase II: Data Pipeline (Semanas 3-4) 17](#_heading=h.xre0n7q5s2c5)

[Fase III: IA/ML (Semanas 5-6) 17](#_heading=h.1y21nh4ysez0)

[Fase IV: Backend API (Semanas 7-8) 17](#_heading=h.4vn6wqvcwaer)

[Fase V: Frontend (Semanas 8-9) 17](#_heading=h.yb3gu1jgrur)

[Fase VI: Testing y Deploy (Semanas 9-10) 18](#_heading=h.mtjj7fcpf0aa)

[10. Control de Cambios 18](#_heading=h.jpisx6br8yny)

[11. Aprobaciones 19](#_heading=h.upl3zafn2zn6)

# **ACTA DE CONSTITUCIÓN DEL PROYECTO INTEGRADOR (PROJECT CHARPET)**

## **1. Información general del Proyecto Integrador**

### **1.1. Nombre del Proyecto Integrador**

ASISTENTE INTELIGENTE DE RUTAS DE APRENDIZAJE BASADO EN IA Y WEB SCRAPING “LearnIA”

### **1.2. Promotor del proyecto Integrador**

Ingeniera Danith Solórzano Escobar

### **1.3. Patrocinador del Proyecto Integrador**

Universidad Pontificia Bolivariana

### **1.4. Definición del Product Owner**

Raúl Ferney Lozano Navarro

### **1.5. Definición del Scrum Master**

Karen Nicol Dayanna Lizarazo Ortiz

### **1.6. Definición del Scrum Team**

El equipo está conformado por María Angelica Parra Grazt, Raúl Ferney Lozano Bottia, Karen Nicol Dayanna Lizarazo Ortiz, Jerson Arley Porras Lozano y Juan Andres Contreras Hernandez.

### **1.7. Interesados (Stakeholders)**

Ingeniera Danith Solórzano Escobar

### **1.8. Socio(s)**

Ingeniera Danith Solórzano Escobar

### **1.9. Fecha de inicio esperada**

22 de julio de 2025

### **1.10. Fecha de finalización**

30 de septiembre de 2025

## **2. Detalles del Proyecto Integrador**

### **2.1. Declaración de la visión del Proyecto Integrador**

Desarrollar una plataforma web inteligente que utiliza técnicas de web scraping, inteligencia artificial y arquitectura cloud nativa para generar rutas de aprendizaje personalizadas, automatizando la curación de contenido educativo de múltiples plataformas y proporcionando recomendaciones basadas en IA según las necesidades específicas de cada usuario.

### **2.2. Planteamiento del problema**

En la era digital actual existe una **sobrecarga de información educativa** dispersa en múltiples plataformas, lo que dificulta a los estudiantes crear rutas de aprendizaje coherentes y eficientes. Los estudiantes pierden tiempo significativo navegando entre múltiples plataformas educativas sin una guía clara sobre la secuencia óptima de aprendizaje, dependencias entre temas, y recursos más adecuados para su nivel y objetivos específicos.

**Problemas identificados:**

* Falta de personalización en las recomendaciones educativas
* Duplicación de esfuerzos al estudiar contenido redundante
* Dificultad para identificar prerrequisitos y dependencias entre temas
* Ausencia de una visión integrada del ecosistema educativo digital

### **2.3. Misión**

Democratizar el acceso a rutas de aprendizaje personalizadas y de calidad mediante la automatización inteligente de la curación de contenido educativo, proporcionando a los usuarios una plataforma unificada que optimice su tiempo de aprendizaje y maximice su efectividad educativa.

### **2.4. Visión**

Ser la plataforma de referencia para la generación automática de rutas de aprendizaje personalizadas, transformando la manera en que las personas acceden y consumen contenido educativo digital a través de inteligencia artificial y tecnologías cloud nativas.

### **2.5. Alcance**

#### **Inclusiones (Qué SÍ está incluido):**

* Web scraping automatizado de 4+ plataformas educativas (Coursera, edX, Udemy, Khan Academy)
* Procesamiento de 10,000+ recursos educativos
* Sistema de IA para clasificación y recomendaciones
* Generación de rutas personalizadas basadas en objetivos del usuario
* Aplicación web completa con interfaz intuitiva
* Despliegue cloud nativo en AWS con arquitectura serverless
* Base de datos optimizada en MongoDB Atlas
* API REST completa para integración frontend-backend

#### **Exclusiones (Qué NO está incluido):**

* Aplicación móvil nativa (iOS/Android)
* Sistema de pagos o monetización
* Integración con LMS existentes (Moodle, Canvas)
* Generación automática de contenido educativo
* Sistema de certificaciones o acreditaciones
* Funcionalidades sociales (chat, foros, colaboración)
* Soporte multiidioma (solo español/inglés)
* Scraping de plataformas premium que requieran suscripción

### **2.6. Objetivos**

#### **Objetivo General:**

Desarrollar y desplegar una plataforma web inteligente basada en inteligencia artificial que genere rutas de aprendizaje personalizadas mediante la extracción, procesamiento y análisis automatizado de recursos educativos de múltiples fuentes digitales.

#### **Objetivos Específicos:**

**OE1: Implementación de Web Scraping Automatizado**

* Desarrollar scrapers para al menos 4 plataformas educativas principales
* Extraer y almacenar información de mínimo 10,000 recursos educativos
* Implementar mecanismos de actualización automática cada 6 horas

**OE2: Desarrollo de Pipeline de Inteligencia Artificial**

* Implementar sistema de procesamiento de lenguaje natural (NLP)
* Desarrollar modelo de clasificación automática de contenido educativo
* Crear sistema de recomendaciones basado en similarity learning
* Generar embeddings semánticos para búsqueda inteligente

**OE3: Arquitectura Cloud Nativa Escalable**

* Desplegar infraestructura completa en AWS utilizando servicios gratuitos
* Implementar arquitectura serverless con auto-scaling automático
* Configurar base de datos MongoDB Atlas para almacenamiento optimizado
* Establecer pipeline CI/CD para deployments automatizados

**OE4: Interfaz de Usuario Intuitiva**

* Desarrollar aplicación web responsiva
* Implementar visualización interactiva de rutas de aprendizaje
* Crear sistema de tracking de progreso del usuario
* Diseñar UX optimizada para diferentes dispositivos

**OE5: Validación y Optimización del Sistema**

* Realizar pruebas de performance y carga
* Validar precisión de recomendaciones con usuarios reales
* Implementar sistema de monitoreo y métricas
* Optimizar costos y recursos del sistema

### 

### **2.7. Beneficios esperados**

1. **Automatización de la curación de contenido** educativo, reduciendo el tiempo de búsqueda
2. **Personalización basada en IA** para optimizar el tiempo de aprendizaje individual
3. **Integración multiplataforma** ofreciendo una vista unificada del ecosistema educativo
4. **Democratización del acceso** a rutas de aprendizaje de calidad sin costo para el usuario
5. **Escalabilidad técnica** para soportar usuarios simultáneamente
6. **Contribución académica** al campo de EdTech y AI aplicada

## 

## 

## 

## **3. Especificación de requerimientos**

### **Requerimientos Funcionales:**

* **RF01:** El sistema debe extraer automáticamente información de al menos 4 plataformas educativas
* **RF02:** El sistema debe procesar y almacenar mínimo 10,000 recursos educativos
* **RF03:** El sistema debe generar recomendaciones personalizadas basadas en IA
* **RF04:** El sistema debe permitir búsqueda semántica avanzada
* **RF05:** El sistema debe visualizar rutas de aprendizaje de forma interactiva
* **RF06:** El sistema debe funcionar como aplicación web responsiva
* **RF07:** El sistema debe actualizar datos automáticamente cada 6 horas
* RF08: El sistema debe tener una interfaz web intuitiva y estética.
* RF09: El sistema debe permitir registrar sus intereses y rutas de aprendizaje
* RF10: El sistema debe incluir autenticación y autorización (incluye registro, login, recuperación de contraseña)

### **Requerimientos No Funcionales:**

* **RNF01:** Tiempo de respuesta menor a 2 segundos para generación de rutas
* **RNF02:** Disponibilidad del sistema mayor al 99%
* **RNF03:** Capacidad para soportar 1000+ usuarios simultáneos
* **RNF04:** Compatibilidad con navegadores modernos (Chrome, Firefox, Safari, Edge)
* **RNF05:** Optimización para dispositivos móviles y tablets
* **RNF06:** Cifrado de datos en tránsito y en reposo

### 

### 

### **Requerimientos Técnicos:**

* **RT01:** Despliegue en AWS utilizando servicios de capa gratuita
* **RT02:** Base de datos MongoDB Atlas M0 (gratuita)
* **RT03:** Frontend desarrollado en React.js
* **RT04:** Backend con Python y FastAPI
* **RT05:** Implementación de CI/CD con GitHub Actions
* **RT06:** Monitoreo con CloudWatch

## **4. Identificación de riesgos**

| **Riesgo** | **Probabilidad** | **Impacto** | **Estrategia de Mitigación** |
| --- | --- | --- | --- |
| **Rate limiting de APIs externas** | Alta | Medio | Implementar delays inteligentes y rotación de IPs |
| **Cambios en estructura de sitios web** | Media | Alto | Desarrollar scrapers modulares y fácil mantenimiento |
| **Límites de AWS Free Tier** | Baja | Alto | Monitoreo continuo de uso y optimización proactiva |
| **Performance de modelos ML** | Media | Medio | Testing temprano y optimización incremental |
| **Complejidad de integración** | Media | Alto | Desarrollo incremental con testing continuo |
| **Disponibilidad de datos** | Baja | Alto | Múltiples fuentes de datos y backup strategies |
| **Cumplimiento legal (ToS)** | Media | Alto | Revisión legal de términos y scraping ético |
| **Retraso en el desarrollo** | Media | Alto | Planificación detallada y Daily Scrum para detectar bloqueos a tiempo |
| **Pérdida de datos** | Media | Alta | Implementación de Backups diarios. |
| **Compatibilidad entre sistemas o software** | Baja | Media | Estandarizar versiones y uso de virtualización y pruebas de integración continua |

## **Análisis DOFA (Debilidades, Oportunidades, Fortalezas, Amenazas)**

### **Fortalezas**

* **F1: Expertise técnico diversificado** - Equipo multidisciplinario con conocimientos en IA, desarrollo web, DevOps y UX/UI
* **F2: Arquitectura cloud nativa escalable** - Implementación en AWS con servicios serverless que garantizan escalabilidad automática
* **F3: Innovación tecnológica** - Combinación única de web scraping, IA y recomendaciones personalizadas en el sector educativo
* **F4: Metodología ágil estructurada** - Implementación de Scrum con roles definidos y sprints planificados

**Oportunidades**

* **O1: Transformación digital en educación** - Cambio paradigmático hacia metodologías de aprendizaje híbrido y personalizado que permite al equipo adquirir experiencia en tecnologías disruptivas del sector educativo
* **O2: Demanda de personalización educativa** - Tendencia creciente hacia aprendizaje adaptativo y rutas personalizadas que brinda al equipo oportunidad de especializarse en sistemas de recomendación e IA aplicada
* **O3: Ecosistema de datos educativos abiertos** - Abundancia de contenido educativo disponible en plataformas públicas que facilita el aprendizaje práctico en web scraping, ETL y gestión de big data
* **O4: Avances en IA**  - Disponibilidad de herramientas y modelos pre-entrenados que permite al equipo desarrollar competencias técnicas en machine learning sin partir desde cero

**Debilidades**

* **D1: Experiencia limitada en proyectos de gran escala** - Equipo conformado por estudiantes sin experiencia profesional extensa
* **D2: Restricciones presupuestarias** - Dependencia de servicios gratuitos que pueden limitar funcionalidades avanzadas
* **D3: Tiempo de desarrollo limitado** - Período de 10 semanas puede ser insuficiente para implementar todas las funcionalidades deseadas
* **D4: Complejidad técnica del scraping** - Riesgo de cambios en APIs y estructuras de sitios web que afecten la extracción de datos
* **D5: Dependencia de servicios externos** - Vulnerabilidad ante cambios en políticas de las plataformas educativas objetivo

**Amenazas**

* **A1: Cambios en términos de servicio** - Plataformas educativas pueden prohibir o restringir el web scraping
* **A2: Competencia establecida** - Existencia de players consolidados como Coursera, edX con recursos superiores
* **A3: Límites técnicos de AWS Free Tier** - Posible exceso de límites gratuitos que generen costos inesperados
* **A4: Velocidad de cambio tecnológico** - Obsolescencia rápida de tecnologías y modelos de IA utilizados.

## **5. Criterios de aceptación del Proyecto Integrador**

### **Criterios Funcionales:**

* 10,000+ recursos educativos extraídos y procesados exitosamente
* Sistema de IA funcional generando recomendaciones relevantes (>80% precisión)
* Aplicación web desplegada y accesible
* Tiempo de respuesta < 1 minuto para generación de rutas

### **Criterios Técnicos:**

* Arquitectura serverless completamente funcional en AWS
* Base de datos optimizada con queries
* API coverage > 80% con documentación completa
* Frontend responsive funcionando en mobile y desktop
* Pipeline CI/CD automatizado y funcional

### **Criterios de Calidad:**

* Código documentado con coverage > 80%
* Usability testing con feedback positivo (>4/5 rating)
* Security assessment sin vulnerabilidades críticas
* Documentación completa de usuario y técnica

### 

### 

### **5.1. Duración del Sprint**

**2 semanas por Sprint**

### **5.2. Criterios de terminado (Done)**

1. Funcionalidad desarrollada y testeada
2. Code review completado
3. Documentación actualizada
4. Deploy exitoso en ambiente de staging
5. Pruebas de aceptación pasadas
6. Sin bugs críticos pendientes

## **6. Cronograma de lanzamientos**

| **Sprint** | **Duración** | **Fechas** | **Entregables Principales** |
| --- | --- | --- | --- |
| **Sprint 1** | 3 semanas |  | Arquitectura definida, infraestructura base, scrapers funcionales |
| **Sprint 2** | 2 semanas |  | Pipeline IA/ML, APIs backend, 10,000 datos procesados |
| **Sprint 3** | 2 semanas |  | Frontend completo, testing, deploy final |

### **Hitos Importantes:**

1. **Semana 2:** Primer scraper funcional con 1,000 datos
2. **Semana 4:** Infraestructura AWS configurada
3. **Semana 6:** Sistema de IA generando primeras recomendaciones
4. **Semana 8:** API backend completa
5. **Semana 10:** Sistema completo desplegado en producción

## **7. Documentación de soporte**

### **Documentación Técnica:**

* Manual de instalación y configuración
* Documentación de APIs
* Arquitectura del sistema
* Manual de operaciones y mantenimiento

### **Documentación de Usuario:**

* Manual de usuario final
* Guías de inicio rápido
* FAQ y troubleshooting
* Videos tutoriales

### 

### 

### 

### **Documentación de Proyecto:**

1. Acta de constitución (este documento)
2. Plan de gestión de riesgos
3. Reportes de avance por Sprint
4. Informe final del proyecto

## **8. Recursos Requeridos**

### **8.1 Recursos Humanos**

| **Rol** | **Dedicación** | **Responsabilidades** |
| --- | --- | --- |
| **Project Manager/Scrum Master** | 100% | Coordinación, planificación, seguimiento |
| **Full Stack Developer** | 100% | Backend, frontend, integración |
| **DevOps/Cloud Engineer** | 100% | Infraestructura, CI/CD, monitoreo |
| **Data Scientist** | 50% | Modelos ML, análisis de datos |

### **8.2 Recursos Tecnológicos**

| **Recurso** | **Costo** | **Justificación** |
| --- | --- | --- |
| **AWS Free Tier** | $0/mes | Infraestructura cloud completa |
| **MongoDB Atlas M0** | $0/mes | Base de datos managed |
| **GitHub** | $0/mes | Control de versiones y CI/CD |
| **Herramientas desarrollo** | $0/mes | VS Code, Postman, etc. |

## 

## **9. Estructura de Desglose del Trabajo (EDT)**

### **Fase I: Fundación (Semanas 1-2)**

* 1.1 Diseño de arquitectura cloud híbrida
* 1.2 Setup de infraestructura AWS
* 1.3 Configuración de MongoDB Atlas
* 1.4 Setup de repositorio y CI/CD

### **Fase II: Data Pipeline (Semanas 3-4)**

* 2.1 Desarrollo de web scrapers
* 2.2 Pipeline de procesamiento de datos
* 2.3 Integración con S3 y Lambda
* 2.4 Población inicial de base de datos

### **Fase III: IA/ML (Semanas 5-6)**

* 3.1 Implementación de modelos NLP
* 3.2 Sistema de clasificación automática
* 3.3 Motor de recomendaciones
* 3.4 Generación de embeddings semánticos

### **Fase IV: Backend API (Semanas 7-8)**

* 4.1 Desarrollo de endpoints REST
* 4.2 Integración con Lambda functions
* 4.3 Sistema de autenticación
* 4.4 Optimización de performance

### 

### **Fase V: Frontend (Semanas 8-9)**

* 5.1 Desarrollo de componentes
* 5.2 Visualizaciones interactivas
* 5.3 UX/UI responsivo
* 5.4 Integración con backend

### 

### **Fase VI: Testing y Deploy (Semanas 9-10)**

* 6.1 Testing integral del sistema
* 6.2 Deploy a producción
* 6.3 Monitoreo y optimización
* 6.4 Documentación final

## **10. Control de Cambios**

| **Versión** | **Fecha** | **Descripción del Cambio** | **Autor** |
| --- | --- | --- | --- |
| 1.0 | 2025-07-25 | Versión inicial del acta de constitución | Grupo PI 2 |

## 

## 

## 

## 

## 

## **11. Aprobaciones**

| **Rol** | **Nombre** | **Firma** | **Fecha** |
| --- | --- | --- | --- |
| **Stakeholder** | Danith Solorzano Escobar | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_ |
| **Product Owner** | [Raúl Ferney Lozano Navarro] |  |  |
| **Scrum Master**  **(Actual)** | [Karen Nicol Dayanna Lizarazo Ortiz] |  | \_\_\_\_\_\_\_ |